

Análise de atividade de instituições de ensino superior em redes sociais

Ana Catarina Marinho Ribeiro

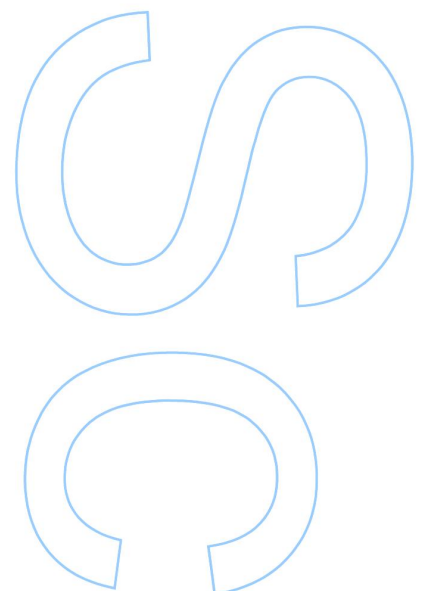
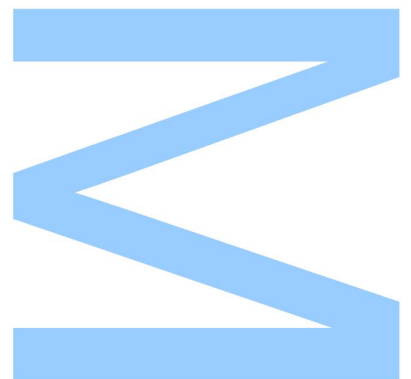
Mestrado Integrado em Engenharia de Redes e Sistemas Informáticos
FCUP-DCC
2017

Orientador

Álvaro Figueira, FCUP-DCC

Coorientador

Luciana Oliveira, ISCAP

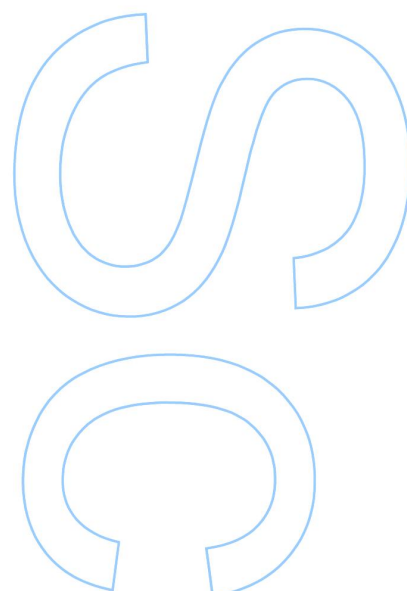
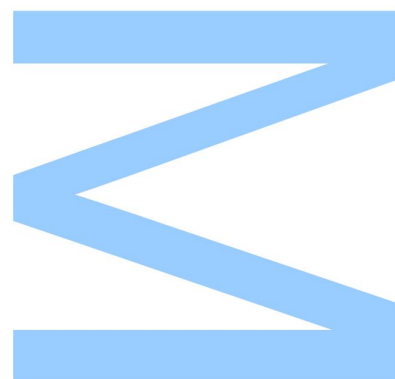




Todas as correções determinadas
pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____



Abstract

Over the last few decades there has been an enormous technological evolution in what concerns web services. Not surprisingly, the importance of the Internet in the ten university institutions that have topped the world ranking has been verified over time, particularly in their use of social networks, namely with Facebook, and in the interactivity between them and their target audience.

In this context, arises a need to create a process to analyze the amount of generated information from the published messages by the institutions in social networks, and the response these messages have in order to compare effort with response.

However, in a time when data is always in high quantities (big data) it is practically impossible to undertake this process solely by manual verification and analysis.

In this context, we propose an automatic procedure to assess the behavior and the individual performance of each institution, followed by an analysis of the posts where the topics contained in the messages were identified. A statistical study was also conducted on the frequency and intensity of publications by universities, which included an analysis of the number of responses to publications over time and on a quarterly basis and a similar analysis of the number of posts. Finally, the content of the texts themselves was analyzed according to the topics covered in the messages.

This process allowed us to identify the most efficient institutions in social networks and the institutions that have a more active community. We have also been able to understand which topics are most common in each university's publications and relate them to the corresponding response from their publics.

Resumo

Ao longo das últimas décadas registou-se uma enorme evolução tecnológica no que diz respeito aos serviços web, nomeadamente na importância da *Internet* para as dez universidades mais bem classificadas no *ranking* mundial, que foram analisadas ao longo do tempo, relativamente à utilização que fazem das redes sociais, particularmente o *Facebook*, e à interatividade gerada entre estas e os seus públicos.

Neste contexto, torna-se premente a criação de um processo que permita analisar a quantidade de informação gerada a partir das mensagens publicadas pelas instituições nas redes sociais, bem como o envolvimento gerado com os fãs, para que seja possível comparar o esforço e a resposta obtida em cada um dos casos.

No entanto, numa altura em que os dados abundam em elevadíssimas quantidades (*big data*) é praticamente impossível realizar este processo recorrendo exclusivamente a análise e verificação manual.

Neste cenário, propomos um processo automático de avaliação de comportamento e de performance individual para cada instituição, seguido de uma análise das mensagens publicadas por estas, onde foram identificados os respetivos tópicos. Foi ainda conduzido um estudo estatístico acerca da frequência e intensidade de publicação de mensagens por parte das universidades, que inclui uma análise trimestral do número de mensagens e respetivas respostas. Finalmente, foi realizada uma análise de conteúdo dos textos das mensagens, de acordo com os principais temas detetados nestas.

Este processo permitiu-nos identificar quais foram as instituições com atividade mais eficiente em redes sociais ao longo do período analisado, bem como aquelas que possuem comunidades mais ativas. Este trabalho permitiu-nos ainda fazer o levantamento dos temas mais frequentemente abordados por cada instituição e relacioná-los com a quantidade de resposta que estes geram junto das respetivas comunidades de seguidores.

Agradecimentos

Gostaria de agradecer a todas as pessoas que, ao longo deste ano de trabalho, me ajudaram e tornaram possível concluir uma grande etapa da minha vida, pois sem a ajuda de todos, tudo isto não seria possível. Em especial ao professor Álvaro Figueira e Luciana Oliveira que, graças aos seus conselhos e dicas, possibilitaram a conclusão deste projeto.

Aos meus pais que lutaram para que pudesse concluir esta etapa durante estes anos, que mesmo nos meus piores momentos acreditaram sempre em mim e estiveram presentes para me apoiar.

A todos os meus amigos que estiveram presentes, quer nos momentos de estudo, quer nos de lazer, mas que também foram um grande apoio para a conclusão desta etapa.

E a todos que de alguma forma estiveram envolvidos em todo este processo, um muito obrigada.

Conteúdo

Abstract	i
Resumo	iii
Agradecimentos	v
Conteúdo	ix
Lista de Tabelas	xi
Lista de Figuras	xv
Lista de Blocos de Código	xvii
Acrónimos	xix
1 Introdução	1
1.1 Contexto e Motivação	1
1.2 Objetivos	2
1.3 Etapas do projeto	3
1.3.1 Identificação da rede social selecionada e das instituições de ensino superior de topo	3
1.3.2 Recolha de informação das instituições de ensino superior	5
1.3.3 Recolha de <i>posts</i> das instituições educacionais e tratamento de dados . . .	5
1.3.4 Utilização dos métodos de <i>Data Mining</i>	5

1.3.5	Deteção de tópicos	5
1.4	Planificação do projeto	6
1.5	Estrutura da tese	6
2	Conceitos Base e Trabalho Relacionado	9
2.1	Conceitos Base	9
2.1.1	<i>Text Mining</i>	9
2.1.2	Classificadores	10
2.1.3	Seleção de informação	12
2.1.4	Métodos de avaliação	12
2.2	Recursos Utilizados	14
2.2.1	Linguagens de Programação	14
2.2.2	Tecnologias	14
2.3	Projetos relacionados	15
2.4	Problemas esperados	16
3	Análise exploratória	17
3.1	Recolha e tratamento dos dados	17
3.1.1	Configuração de <i>API</i> do <i>Facebook</i>	17
3.1.2	Extração dos dados	18
3.2	Tratamento de dados	19
3.3	Análise dos dados recolhidos	20
3.4	Conclusão	26
4	Análise de comportamento e desempenho	27
4.1	Evolução do comportamento ao longo do tempo	27
4.1.1	Atividade no <i>Facebook</i> ao longo do tempo	28
4.1.2	Interações ao longo do tempo	34
4.2	Análise de desempenho global por trimestres	40

5	Análise de conteúdo	45
5.1	Análise da importância das palavras para o <i>post</i>	45
5.2	Análise dos tópicos das mensagens	51
5.3	Conclusões	55
5.3.1	Importância das palavras	55
5.3.2	Tópicos das mensagens	55
6	Conclusões	57
6.1	Resumo do Trabalho	57
6.2	Conclusões Gerais	57
6.3	Trabalho Futuro	58
	Bibliografia	59
A	Análise dos tópicos contidos nas mensagens	61
B	Calendários escolares	67

Lista de Tabelas

3.1	Síntese dados numéricos	19
3.2	Número de palavras por Instituição	22
3.3	Número de publicações por dia	22

Lista de Figuras

1.1	Planificação	6
2.1	Confusion matrix	13
3.1	Criação da <i>APP</i> para o sistema	17
3.2	Permissões da <i>API</i>	18
3.3	Número de <i>posts</i> por Instituição	20
3.4	Número de interações por Instituição	21
3.5	Número de <i>posts</i> por meses do ano	23
4.1	Agrupamento de <i>posts</i> para <i>Harvard</i>	28
4.2	Agrupamento de <i>posts</i> para <i>Stanford</i>	29
4.3	Agrupamento de <i>posts</i> para <i>MIT</i>	30
4.4	Agrupamento de <i>posts</i> para <i>Cambridge</i>	30
4.5	Agrupamento de <i>posts</i> para <i>Oxford</i>	31
4.6	Agrupamento de <i>posts</i> para <i>Columbia</i>	31
4.7	Agrupamento de <i>posts</i> para <i>Berkeley</i>	32
4.8	Agrupamento de <i>posts</i> para <i>Chicago</i>	33
4.9	Agrupamento de <i>posts</i> para <i>Princeton</i>	33
4.10	Agrupamento de <i>posts</i> para <i>Yale</i>	34
4.11	Agrupamento do número de respostas para <i>Harvard</i>	35
4.12	Agrupamento do número de respostas para <i>Stanford</i>	35

4.13	Agrupamento do número de respostas para <i>MIT</i>	36
4.14	Agrupamento do número de respostas para <i>Cambridge</i>	36
4.15	Agrupamento do número de respostas para <i>Oxford</i>	37
4.16	Agrupamento do número de respostas para <i>Columbia</i>	38
4.17	Agrupamento do número de respostas para <i>Berkeley</i>	38
4.18	Agrupamento do número de respostas para <i>Chicago</i>	39
4.19	Agrupamento do número de respostas para <i>Princeton</i>	39
4.20	Agrupamento do número de respostas para <i>Yale</i>	40
4.21	Desempenho global no primeiro trimestre do ano letivo	41
4.22	Desempenho global no segundo trimestre do ano letivo	41
4.23	Desempenho global no terceiro trimestre do ano letivo	42
4.24	Desempenho global no quarto trimestre do ano letivo	42
5.1	Palavras importantes utilizadas por Harvard University	46
5.2	Palavras importantes utilizadas por Stanford University	46
5.3	Palavras importantes utilizadas por Massachusetts Institute of Technology	47
5.4	Palavras importantes utilizadas por University of Cambridge	47
5.5	Palavras importantes utilizadas por University of Oxford	48
5.6	Palavras importantes utilizadas por Columbia University	48
5.7	Palavras importantes utilizadas por University of California, Berkeley	49
5.8	Palavras importantes utilizadas por University of Chicago	49
5.9	Palavras importantes utilizadas por Princeton University	50
5.10	Palavras importantes utilizadas por Yale University	50
5.11	Agrupamento das instituições <i>Harvard</i> e <i>Yale</i>	51
5.12	Agrupamento das instituições <i>Cambridge</i> , <i>MIT</i> e <i>Chicago</i>	52
5.13	Agrupamento das instituições <i>Columbia</i> e <i>Oxford</i>	53
5.14	Agrupamento das instituições <i>Princeton</i> , <i>Berkeley</i> e <i>Stanford</i>	54
A.1	Tópicos obtidos em <i>Harvard</i>	61

A.2	Tópicos obtidos em <i>Stanford</i>	62
A.3	Tópicos obtidos em <i>MIT</i>	62
A.4	Tópicos obtidos em <i>Cambridge</i>	63
A.5	Tópicos obtidos em <i>Oxford</i>	63
A.6	Tópicos obtidos em <i>Columbia</i>	64
A.7	Tópicos obtidos em <i>Berkeley</i>	64
A.8	Tópicos obtidos em <i>Chicago</i>	65
A.9	Tópicos obtidos em <i>Princeton</i>	65
A.10	Tópicos obtidos em <i>Yale</i>	66

Lista de Blocos de Código

Acrónimos

CRAN	Cromprehensive R Archive Network	NLP	Neuro-Linguistic Programming
CU	Columbia University	Oxon	University of Oxford
CWUR	Center for World University Rankings	PU	Princeton University
DCC	Departamento de Ciências de Computadores	ROI	Retorno sobre investimento
FCUP	Faculdade de Ciências da Universidade do Porto	SU	Stanford University
FN	False Negative	SVM	Support Vector Machine
FP	False Positive	TF	Term Frequency
HTTP	Hypertext Tranfer Protocol	TF-IDF	Term Frequency - Inverse Document Frequency
HU	Harvard University	TN	True Negative
IDF	Inverse Document Frequency	TP	True Positive
KNN	K-Nearest Neighbour	UC	University of Cambridge
KPI	Key Performance Indicator	UCB	University California, Berkeley
LDA	Latent Dirichlet Allocation	UCHI	University of Chicago
MIT	Massachusetts Institute of Technology		

Capítulo 1

Introdução

1.1 Contexto e Motivação

A complexidade dos novos problemas que a sociedade enfrenta resulta do seu crescimento e rápida evolução científica e tecnológica, sobretudo registada nas últimas décadas. Em consequência, o desenvolvimento de soluções inovadoras que necessariamente resultem do cruzamento e articulação multidisciplinar de áreas de conhecimento, torna-se vital.

Devido à consolidação da *Internet* como meio preferencial de comunicação na atualidade, as instituições ficam diante do desafio de lidar com o empoderamento dos seus públicos-alvo. A proliferação das redes sociais tem introduzido profundas transformações na sociedade e na comunicação entre pessoas e/ou organizações. Efetivamente as instituições começaram a investir cada vez mais na virtualização e no desenvolvimento da sua presença nestes ambientes digitais, que proporcionam grande interatividade e amplo alcance do público. Nos últimos anos, com o desenvolvimento da *Internet* e o uso frequente das redes sociais, as instituições de ensino superior têm sentido a necessidade de seguir essa tendência mundial de construir presença nas redes sociais, nomeadamente na rede mais utilizada em todo o mundo, o *Facebook*.

Um dos grandes desafios impostos às instituições, com a emergência das redes sociais, reside na necessidade de migrar dos modelos de comunicação assimétricos para uma perspetiva mais simétrica de comunicação, ou seja, mais centrada na opinião e nos interesses das audiências, cujo poder de intervenção é cada vez mais ampliado. A pensar nessa interatividade entre as instituições e o seu público, bem como na facilidade dessa comunicação, a quantidade de informação gerada será bastante elevada para ser tratada manualmente.

De facto, as redes sociais criaram a possibilidade de qualquer pessoa ou instituição distribuir informação de forma veloz e dinâmica. Num simples clique é possível obter informações, publicar conteúdos, comentar e partilhar informação de outros utilizadores. Estas redes assumem-se como estruturas que permitem ligar pessoas ou organizações com interesses idênticos, permitindo uma partilha de experiências e informações entre as partes de modo mais acessível.

No entanto, a construção desta presença, através da criação de páginas e perfis em redes sociais, tem vindo a suceder independentemente do fato de as organizações terem, ou não, previamente definido objetivos específicos para estes canais ou uma estratégia de conteúdo orientada às expectativas de um público-alvo também claramente identificado. Em resultado, são muito poucas as instituições que possuem uma estratégia de conteúdo para redes sociais, embora este seja um tema popular entre os profissionais de *marketing* e uma área crescente de interesse. Neste momento, as instituições encontram-se continuamente a despendar tempo, recursos humanos e tecnologias para assegurar a manutenção da sua presença nas redes sociais e esta alocação de recursos torna cada vez mais premente a necessidade de obter e medir o retorno sobre o investimento nestes meios.

1.2 Objetivos

Este projeto tem como objetivo essencial o desenvolvimento de um sistema automático de classificação de mensagens que são colocadas nas redes sociais pelas instituições do ensino superior internacionais. Este objetivo assenta na premissa de que o conjunto de mensagens publicadas por uma instituição de ensino superior pode ser decomposto em grupos de mensagens análogas (orientadas ao mesmo objetivo genérico).

De acordo com outros estudos [14][20][11], é esperado que pelo menos três áreas surjam ao longo da análise, sendo elas a Educação, Investigação e Relacionamento. No entanto, o que se pretende, ao contrário do estudo Figueira e Oliveira [11], é que as áreas editoriais não sejam prescritas, mas que sejam reveladas através da utilização de um sistema não supervisionado que permita classificar as mensagens publicadas em redes sociais em áreas editoriais que sejam reveladas através da análise do conteúdo dessas mesmas mensagens.

Os objetivos específicos deste trabalho consistem em:

1. Identificar as instituições de ensino superior que serão alvo de estudo.
2. Recolher todas as mensagens publicadas nas redes sociais durante um ano letivo completo.
3. Identificar e classificar os termos predominantes das mensagens recolhidas.
4. Computar o envolvimento dos públicos com os temas predominantes identificados, utilizando KPI com *likes*, *reactions*, *comments* e *shares*.
5. Identificar a geração não-supervisionada de tópicos contidos nas mensagens.
6. Identificar os temas com propensão para gerar mais envolvimento dos públicos a fim de fornecer recomendações para o desenho de estratégias de conteúdo.
7. Identificar as instituições internacionais mais bem posicionadas em termos de eficiência no desenho de conteúdo e envolvimento gerado nos públicos.

1.3 Etapas do projeto

A abordagem proposta é constituída por 4 fases fundamentais.

Numa primeira fase, verificamos, qual a rede social mais utilizada pelas instituições de ensino superior internacional. A identificação das instituições de ensino superior de topo é resultante do seguimento do *ranking* internacional *Center for World University Rankings* (CWUR) [7], para assim ter acesso às melhores e mais qualificadas instituições de ensino superior a nível mundial.

Assim, a segunda fase consiste em reunir toda a informação, relativa a um ano letivo, da página oficial do Facebook de cada instituição, através de um *crawler* desenvolvido com esse propósito, que recolhe todos os *posts*, *comments*, *shares* e *reactions*. Todos esses dados são submetidos a um processo de limpeza dos dados recolhidos, o que facilita o estudo em causa.

A terceira fase da metodologia centra-se na seleção de métodos para a realização de uma análise aos dados, com a utilização de métodos que simplifiquem a compreensão dos dados recolhidos e permita o estudo do comportamento e desempenho das instituições.

Finalmente, a última fase desta metodologia consiste no desenvolvimento de um estudo exploratório para deteção automática de estratégias de conteúdo de redes sociais com base nas *keywords* utilizadas nas mensagens.

1.3.1 Identificação da rede social selecionada e das instituições de ensino superior de topo

Numa primeira etapa, verificamos, através do estudo Figueira, Á. and Oliveira, L., 2016[13], que o *Facebook* é a rede social mais significativa no âmbito das instituições de ensino superior, devido ao elevado número de utilizadores, à distribuição desses utilizadores pela rede mundial, aos vários modos e formas de interagir com o seu público.

Para a realização do presente estudo foi seguido um *ranking*, para assim ter acesso às melhores e mais qualificadas instituições de ensino superior a nível mundial. O *Center for World University Rankings*, mais conhecido por CWUR, publica anualmente a classificação internacional das universidades, em que mede a qualidade na educação e formação dos alunos, o prestígio dos docentes que nelas operam, incluindo a qualidade da investigação por eles produzida [7].

O *ranking* teve origem na Arábia Saudita, em 2012, com o objetivo de classificar as 100 melhores instituições de ensino superior a nível mundial. O seu sucesso foi de tal forma reconhecido, que rapidamente foi expandido para classificar as melhores 1000 instituições de ensino superior internacionais, tornando-se assim no melhor *ranking* académico do mundo.

A construção do *ranking* referido é feita com base em oito objetivos, como se indicam: *Quality of Education*, *Alumni Employment*, *Quality of Faculty*, *Publications*, *Citations*, *Broad Impact*, *Influence* e *Patents*. Cada um destes objetivos tem o seu grau de importância, permitindo assim

criar a classificação das melhores 1000 instituições de ensino superior a nível mundial [7].

Quality of Education é medido pelo número de alunos de uma universidade que ganharam prêmios de renome internacional. Esta relação mede a qualidade da educação de uma instituição com base no futuro académico dos seus alunos. Este objetivo tem uma grande importância no *ranking*, uma vez que o seu peso é de 25%.

Alumni Employment é medida pelo número de alunos de uma dada universidade que ocupam a posição de CEO de grandes empresas de topo em relação ao tamanho da universidade em questão. Tal como o objetivo *supra* descrito, este tem grande importância no *ranking*, pois que a percentagem contabilizada é de 25%

Quality of Faculty, este indicador mede o número de professores de uma instituição que ganharam medalhas e prêmios importantes que incorporam praticamente todas as medalhas de carisma das disciplinas académicas. Os membros do corpo docente são definidos aqui como aqueles que estavam empregados na instituição no momento de ganhar o prémio ou medalha. Este objetivo tem, também, um peso de 25%.

Publications, neste estudo é medido o número de publicações em revistas conceituadas nas áreas das Ciências, Ciências Sociais e Humanidades. Este objetivo tem uma percentagem de 5%, que é relativamente baixa em relação aos mencionados anteriormente.

Broad Impact é medido pelo h-index das universidades, ou seja, o h-index é um número destinado a medir tanto a produtividade como o impacto do trabalho publicado. O peso deste objetivo é de 5% no cálculo final para a realização do *ranking*.

Influence, para este indicador é considerado o número de trabalhos de pesquisa completos que foram publicados em jornais e/ou revistas influentes numa dada área de investigação. Este ponto em questão tem um peso no cálculo final de 5%, igualando-o assim a outros objetivos já vistos.

Patents, este é o último objetivo utilizado para a criação do *ranking*, também com uma percentagem de 5% no cálculo final. É medido pelo número de pedidos feitos para patentes internacionais, ou seja, as instituições podem então ser classificadas com base no número médio de patentes registadas por ano.

Para a realização deste trabalho considerou-se incluir as 10 instituições melhor posicionadas, visto que elas constituem uma análise representativa das melhores universidades a nível mundial. Por conseguinte, o conjunto de instituições que são objeto de estudo, são: *Harvard University*, *Stanford University*, *Massachusetts Institute of Technology*, *University of Cambridge*, *University of Oxford*, *Columbia University*, *University of California, Berkeley*, *University of Chicago*, *Princeton University* e *Yale University*[7].

1.3.2 Recolha de informação das instituições de ensino superior

Esta etapa destina-se à recolha de informação relevante das instituições. Através das páginas *web* recolhem-se os calendários escolares que contribuem tanto para a recolha de informação, uma vez que esses dados indicam o início e o fim do ano letivo de cada instituição, como para a análise da informação gerada.

1.3.3 Recolha de *posts* das instituições educacionais e tratamento de dados

Nesta etapa é desenvolvido um *crawler* que, para o propósito desta investigação, recolhe todos os *posts*, número de *comments*, *shares* e *reactions* da página oficial de cada instituição de ensino superior envolvida no presente estudo. Para obter uma análise abrangente, os dados recolhidos estão incluídos num período "representativo", selecionado de forma que contenha um ano letivo de todas as instituições.

Uma vez recolhida toda a informação necessária, realiza-se uma análise léxica, para remover todos os *emojis*, acentos, pontuações, espaços em excesso e converter todas as letras para minúsculas. De seguida, remove-se todos os *stop-words*, que são todos os termos considerados irrelevantes e que apenas têm um papel funcional ao longo do *post*.

Assim que o processo de recolha e tratamento termine, efetua-se um breve estudo exploratório inicial ao conjunto de dados recolhidos das instituições.

1.3.4 Utilização dos métodos de *Data Mining*

Após o tratamento de todos os dados, são identificados os dados de carácter ruidoso. Para isso são usadas técnicas como *binning*, que suavizam o ruído existente, agrupando-os em intervalos. Estes agrupamentos são executados tendo em conta o número de publicações divulgadas e as respostas obtidas a esses *posts*, toda esta análise é elaborada com recurso aos calendários escolares e às páginas oficiais de cada instituição. De seguida, efetua-se um estudo que possibilita a análise do comportamento das instituições, criando agrupamentos que incluam o número de *posts* e a quantidade de interação entre os seguidores e a página oficial da entidade, soma dos *likes*, *comments* e *reactions*.

1.3.5 Detecção de tópicos

A partir deste momento, todo o estudo é redirecionado para a análise das mensagens divulgadas pelas instituições nas suas páginas do *Facebook*. Esta etapa tem o propósito de identificar as palavras e os diferentes tópicos que mais se destacam nas publicações realizadas. Para este efeito, os classificadores de DM são adaptados para possibilitar a obtenção de resultados.

1.4 Planificação do projeto

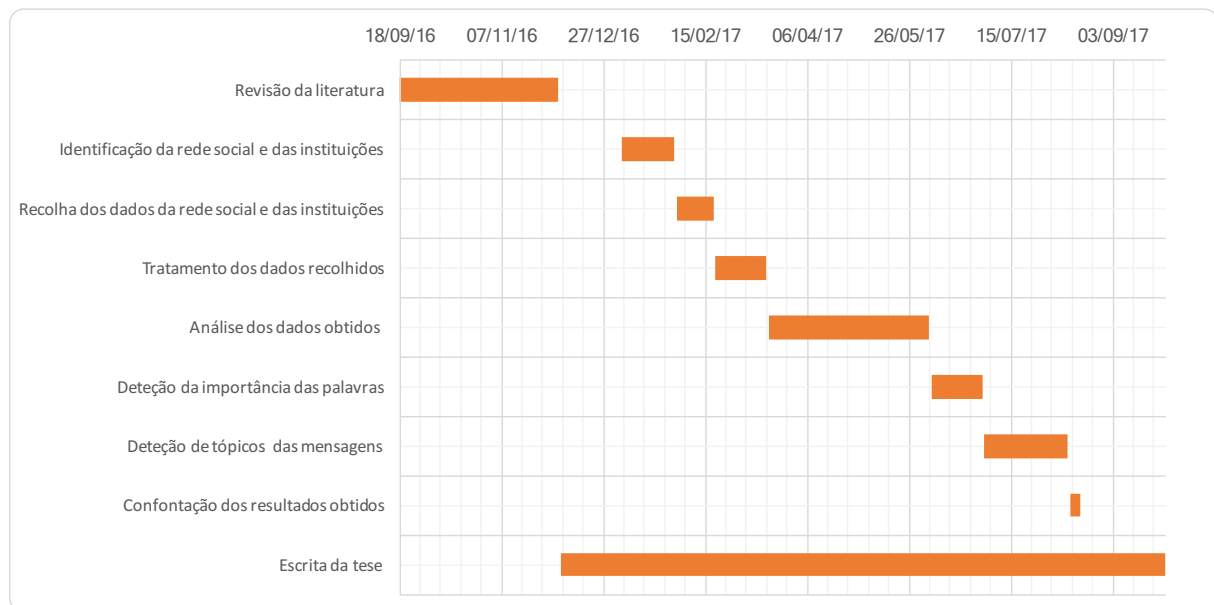


Figura 1.1: Planificação

Como mencionado em cima, a realização deste projeto divide-se em várias etapas, de forma a permitir uma boa gestão das tarefas delineadas para dar resposta aos objetivos do trabalho, o desenho desta planificação é realizado de modo a que as tarefas subsequentes sejam corretamente distribuídas num espaço temporal que permite contar com o *input* das tarefas precedentes. Para isso, recorre-se ao diagrama de *Gantt*, para sistematizar as várias fases a serem realizadas, qual a sua duração e de que forma as tarefas previstas se encontram integradas e articuladas num único processo global. Na figura 1.1 é ilustrado o planeamento para este projeto. A primeira fase, que se baseia no estudo e na análise das necessidades e das ferramentas posteriormente consideradas na implementação, de seguida temos as etapas relacionadas com o desenvolvimento dos métodos de DM e análise dos resultados obtidos. Por fim, a etapa de redação da tese que acompanha todas as restantes etapas de desenvolvimento do projeto.

1.5 Estrutura da tese

Tendo este documento o objetivo de detalhar o estudo realizado sobre estratégias de conteúdo nas redes sociais pelas instituições de topo de ensino superior, apresentam-se as conclusões obtidas que foram sustentadas em várias etapas realizadas ao longo do processo.

Para além da introdução, são apresentadas todas as etapas utilizadas para a realização do processo e o planeamento para o estudo desenvolvido. Esta tese é composta por mais cinco capítulos.

No capítulo 2, é realizada uma apresentação dos conceitos base e os recursos utilizados para

o desenvolvimento do projeto em causa. Neste capítulo são também apresentados os projetos a que este está relacionado e os problemas esperados na realização do mesmo.

A fase de recolha e tratamento dos dados é descrita ao longo do capítulo 3, sendo apresentado os passos necessários utilizados para estabelecer a conexão com a *API* do *Facebook* e assim realizar a recolha dos dados pretendidos tratamento dos mesmos. Finalizada esta fase, é realizado um estudo exploratório inicial.

No capítulo 4 é demonstrado o estudo de desempenho por universidade, onde se examina a quantidade de publicações realizadas e a atividade existente ao longo do tempo entre a comunidade e a rede social. Por fim, é estudada a performance das instituições por trimestre.

No capítulo 5 é descrito todo o processo de análise de conteúdo existente nas publicações, em primeiro lugar analisa-se a importância das palavras utilizadas e em seguida, explora-se os tópicos utilizados nas mensagens.

Por fim, são apresentadas as conclusões retiradas com todo este estudo e realizada uma análise ao trabalho futuro possível a ser desenvolvido para completar este trabalho.

Capítulo 2

Conceitos Base e Trabalho Relacionado

2.1 Conceitos Base

O principal objetivo desta secção consiste na apresentação das técnicas existentes na literatura e que são úteis de forma a alcançar os objetivos inicialmente definidos e tentar encontrar outras propostas para os problemas já identificados.

Para ser possível uma melhor resolução do problema proposto, recorre-se ao uso do *Text Mining*, em especial às técnicas de **classificação de texto**, uma vez que o que se pretende é classificar dados retirados do Facebook.

2.1.1 *Text Mining*

Normalmente, é definido como um processo que utiliza procedimentos para organizar, encontrar e descobrir informação em documentos escritos na linguagem natural, pelo facto do *Text Mining* remover informação considerada útil de dados não estruturados e/ou semi-estruturados e por estes serem de difícil tratamento [16].

Text Mining é uma extensão natural do *Data Mining*, mas aplicado a dados textuais, que surgiu da conveniência de encontrar informação nos textos, de forma autónoma. Têm ambos uma grande relevância nas áreas de classificação de texto devido ao crescimento da Internet, que nos dias de hoje faz parte do dia-a-dia de muitas pessoas.

O uso desta técnica do *Text Mining* possibilita a recuperação e extração de informação, classificação de textos, resumos de documentos e possibilita também a realização de análises qualitativas e quantitativas de documentos de texto.

A classificação de texto neste caso, mineração de texto, ou *text mining*, reporta a dificuldade existente em atribuir categorias pré-definidas automaticamente a documentos de texto. De uma forma geral, os trabalhos referentes à classificação de texto pretendem encontrar um tema central para o texto em questão. Para que isso se concretize, é necessário seguir alguns passos [5]:

- Pré-processamento: definido muitas vezes como um processo transformativo, uma vez que transforma o documento inicial de forma a facilitar o processo de classificação. O pré-processamento é aplicado, normalmente, para reduzir o número de termos de um documento de modo a obter uma representação do documento mais adequado para as fases que se seguem. Nesta fase, são realizadas diversas tarefas como:
 - Análise léxica: permite remover toda a pontuação, acentos existentes ao longo do documento e todas as letras do texto em maiúsculas passando para minúsculas.
 - Remoção de *Stop-words*: as *stop-words* são todos os termos considerados irrelevantes que apenas têm um papel funcional ao longo do *post*. Estas palavras derivam muito do idioma analisado, e normalmente são removidas para assim poder melhorar os métodos de processamento. As palavras que são removidas têm de ser palavras com uma semântica fraca e assim só ficam as palavras realmente importantes.[17]
 - *Stemming*: é um processo que reduz palavras derivadas, transformando-as na sua palavra raiz. Se for aplicado o *stemming* nas palavras como: computador, computar e computação estas palavras serão reduzidas a "computa", pois é a sua raiz. Com este processo é possível reduzir plurais e conjugações verbais e, assim, o modelo de aprendizagem consegue classificar um documento corretamente, reduzindo também o número de palavras avaliadas, mas tem a desvantagem de, por vezes, existirem palavras com a mesma raiz mas possuírem significados diferentes[9].
- Classificação: nesta etapa realiza-se a criação de classificadores que proporcionam a atribuição de categorias aos documentos. É indispensável a utilização de vários algoritmos de *Data Mining* mais utilizados para a área de *Text Mining*.

Text Mining é incluído no *package tm* da linguagem R que possui algumas ferramentas a partir das quais é possível classificar documentos automaticamente, ter uma ideia do conteúdo de um dado documento sem ter necessidade de o ler[19].

2.1.2 Classificadores

Existe a possibilidade de se identificarem vários classificadores ou métodos que permitem a resolução de problemas na categorização de texto. Alguns classificadores mais comuns para a área de *Text Mining* são:

- *Support Vector Machines*

Um *Support Vector Machine*, mais conhecido por **SVM** é um sistema de aprendizagem supervisionado baseado no conceito de planos de decisão que definem limites de decisão. Um plano de decisão é aquele que separa um conjunto de objetos com diferentes classes. Os **SVMs** podem ser lineares, que encontra um separador (hiperplano) entre duas classes com uma margem máxima, que separa os exemplos positivos e negativos para cada classe. A

margem é a distância entre o hiperplano e o ponto mais próximo de cada uma das classes, os pontos mais próximos são definidos de **vetores de suporte** [1].

Os **SVMs** não lineares são eficazes na classificação de dados linearmente separáveis ou que tenham uma distribuição aproximadamente linear. Nesses casos, o uso da fronteira curva é o mais adequado para poder separar as classes, assim torna-se possível utilizar diferentes *kernels*.

- *N-Grams*

É um sistema robusto e rápido, com muita utilidade para a classificação de texto. Este sistema é tolerante a erros textuais e baseia-se no cálculo e na comparação de perfis de frequências de N-gramas. Usa-se o sistema para calcular o carácter de *data sets* de treino que representam as várias categorias. Em seguida, é calculado um perfil para um documento específico que deve ser classificado. Finalmente, calcula uma medida de distância entre o perfil do documento e cada um dos perfis da categoria. O sistema seleciona a categoria cujo perfil tem a menor distância para o perfil do documento [2].

Um N-Gram é uma sequência de n palavras consecutivas: unigram para n=1, bigram para n=2, trigram para n=3, etc. Por exemplo: “Text Classification” tem 2 palavras, é o chamado bigram.

- *K-Nearest Neighbour* ou **KNN**

Denominado de vizinho mais próximo, é utilizado para classificação, em que permite armazenar todos os casos disponíveis e classifica novos casos com base numa medida de similaridade. É medido por uma função de distância e se $K = 1$, então o caso é simplesmente atribuído à classe de seu vizinho mais próximo.

O valor do k, que corresponde ao número de vizinhos a considerar, a função de semelhança que é usada para encontrar os vizinhos mais próximos e a regra de decisão para que seja possível identificar a classe do documento teste são os principais fatores que podem influenciar o desempenho deste algoritmo. Além do mais, este algoritmo tem um inconveniente significativo, a sua eficácia, uma vez que ele precisa de comparar o documento teste com todas as amostras do conjunto de treino para assim ser possível classificá-lo [1].

- *Decision Trees*

Uma *decision tree*, ou árvore de decisão, é um diagrama de fluxo que oferece a estrutura de uma árvore. Os nós internos que constituem a *decision tree* são utilizados para testes realizados aos atributos e as suas folhas demonstram as categorias.

Algumas das vantagens na utilização deste modelo centram-se na facilidade de interpretação e implementação, no treino e na classificação que é executada de forma ágil. No entanto, as *decision trees*, como qualquer outro método, apresentam inconvenientes: a sua construção é mais difícil, uma vez que o processo é repetido por todas as subárvores existentes, e é também um modelo instável, ou seja, os pequenos distúrbios no conjunto de treino podem gerar alterações no modelo estudado [15].

2.1.3 Seleção de informação

- Term Frequency - Inverse Document Frequency

Term Frequency - Inverse Document Frequency (TF-IDF) é composto por dois termos: *term frequency* (TF), que calcula a frequência que uma palavra ocorre num documento, esses valores são inseridos numa matriz a que se designa de *document-term matrix*. O segundo termo é *inverse document frequency* (IDF) que diminui o peso para palavras usualmente usadas e aumenta o peso para palavras que não são usadas com tanta frequência num dado documento. Consequentemente, o *TF-IDF* mede a importância de uma palavra para um dado documento ou conjunto de documentos. O *IDF* para qualquer termo é dado por [18]:

$$idf(term) = \ln \frac{N_{documents}}{N_{documents\ containing\ term}} \quad (2.1)$$

- Latent Dirichlet Allocation

Latent Dirichlet Allocation ou **LDA** é um modelo estatístico, que permite descobrir de forma automática tópicos presentes nos documentos em que este está a ser executado, apresentando a probabilidade de surgimento de cada um desses tópicos. As variáveis ocultas são as distribuições de tópicos, cujos parâmetros são dados previamente, enquanto que as variáveis visíveis são os termos de cada um dos documentos. Tal como indica o nome deste modelo, a distribuição utilizada é a distribuição de *Dirichlet*, que consiste em alocar dados dos vários tópicos e, posteriormente, preencher os documentos [4].

2.1.4 Métodos de avaliação

Considerada uma fase crítica do processo de *Data Mining*, esta etapa é necessária para que os especialistas de tratamento de dados possam analisar todos os dados recolhidos ao longo de um estudo/projeto e assim ser possível tirar conclusões mais coerentes. A área de *Data Mining* permite usar várias medidas que simplifica a avaliação dos resultados. As medidas mais comuns e indicadas para analisar resultados de classificação de texto são: correlação de *Pearson*, *recall*, *precision*, *fallout*, *accuracy* e *error*, para auxiliar algumas destas medidas, temos a *Confusion matrix*.

- Correlação de *Pearson*

Uma das medidas mais usadas para quantificar o grau de associação linear entre duas variáveis quantitativas. O coeficiente de correlação de *pearson* (r), calcula-se segundo a equação 2.2 e quantifica a semelhança que existe entre dois vetores numéricos. Este valor varia entre -1 e 1. Uma relação diretamente proporcional ocorre quando o coeficiente é igual a 1, uma relação inversa exata ocorre quando o coeficiente é igual a -1. Quando o coeficiente é igual a 0, significa que não existe relação linear entre dois valores[22].

$$r = \frac{\sum (xi - \bar{x})(yi - \bar{y})}{\sqrt{(\sum (xi - \bar{x})^2)(\sum (yi - \bar{y})^2)}} \quad (2.2)$$

- *Confusion matrix*

Também conhecida por matriz de confusão, esta matriz ilustrada em 2.1, para uma determinada hipótese h , oferece uma medida efetiva do modelo de classificação, ao mostrar o número de classificações corretas versus as classificações preditas para cada classe, sobre um conjunto de exemplos T . O número de acertos, para cada classe, localiza-se na diagonal principal da matriz $M(C_i, C_i)$ e os demais elementos $M(C_i, C_j)$, para $i \neq j$, representam erros na classificação. A matriz de confusão de um classificador ideal possui todos esses elementos iguais a zero, uma vez que ele não comete erros. A *confusion matrix* é constituída por **TP**, que são os *true positive*, ou verdadeiros positivos, os **TN**, mais conhecidos por *true negatives*, que são os verdadeiros negativos. Já os **FP**, *false positive*, são os falsos positivos e, por fim, existem ainda os **FN**, que são os *false negative*, considerados os falso negativos.

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Figura 2.1: Confusion matrix

- *Recall*

Calcula a percentagem de amostras positivas classificadas corretamente sobre o total de amostras positivas. Pode-se concluir que:

$$Recall = \frac{TP}{TP + FN} \quad (2.3)$$

- *Precision*

Calcula a percentagem de amostras positivas classificadas corretamente sobre o total de amostras classificadas como positivas, ou seja:

$$Precision = \frac{TP}{TP + FP} \quad (2.4)$$

- *Fallout*

É usado para medir a quantidade de documentos irrelevantes recuperados. A melhor taxa atingida é 0, isso significa que não existe nenhum documento encontrado que seja relevante. Uma vez obtido o resultado há necessidade de o analisar bem, pois pode significar que não foi recuperado nenhum documento.

- *Accuracy*

Possibilita a obtenção da relação dos documentos classificados corretamente sobre o número total de documentos avaliados.

$$Accuracy = \frac{TP + TN}{P + N} \quad (2.5)$$

- *Error*

Permite obter a relação entre os documentos classificados incorretamente sobre o número total de documentos avaliados, ou seja, calcula a fração das previsões incorretas.

$$Error = 1 - \frac{TP + TN}{P + N} \quad (2.6)$$

2.2 Recursos Utilizados

Para a realização do presente trabalho foram selecionadas como ferramentas a linguagem R e Facebook API, que permitem a implementação de todas as fases propostas.

2.2.1 Linguagens de Programação

A linguagem R é uma linguagem de programação e um ambiente para computação, estatística e gráfica. Existem várias versões para Windows, Unix/Linux e Mac, e é uma linguagem que é executada em diferentes arquiteturas [21]. R possui um vasto conjunto de pacotes, mas também é possível aceder a outros através da rede de distribuição do R, o CRAN (Comprehensive R Archive Network). Esta será a linguagem padrão, uma vez que é ótima para trabalhar com a análise de dados devido à grande dimensão de dados tratados [21].

2.2.2 Tecnologias

A *API do Facebook* é uma ferramenta que possibilita o desenvolvimento de aplicações para aceder aos dados do *Facebook*. A *API* permite, de modo programável, consultar, administrar dados, anúncios e muito mais. Esta informação é recolhida através do *Graph API*, que é representada por intermédio de *nodes*, que são elementos como utilizadores, fotos, páginas ou comentários; *edges*, que simbolizam ligações que associam os elementos *nodes*; e, por último, *fields*, que indicam informações associadas aos elementos, como o aniversário de uma pessoa ou o nome de uma determinada página. A *Graph API* é baseada em HTTP e, por isso, funciona com qualquer linguagem que tenha uma biblioteca HTTP [6].

2.3 Projetos relacionados

No decorrer da fase de pesquisa para a elaboração do projeto, focámo-nos em estudos com objetivos semelhantes ao presente projeto, entre eles destacam-se [11], [13] e [12].

Em ambos os estudos, os autores têm como objetivo analisar as mensagens publicadas em redes sociais das Universidades e Politécnicos Portugueses, tendo concluído que a rede social mais utilizada pelas instituições é o *Facebook*.

Para os estudos analisados, foram utilizados o mesmo número de agentes, que incluíam institutos politécnicos integrados em institutos politécnicos e institutos politécnicos integrados em universidades, começando com 137 agentes e diminuindo, posteriormente, para 94, uma vez que, nesse momento, são foram incluídas instituições que prestam serviços educativos, onde são ignoradas as entidades gestoras de cada instituto e universidade. Nos estudos [11] e [13] foram realizados com ajuda humana, ou seja, antes de a informação ser classificada por algoritmos era categorizada de forma manual para ser possível identificar as categorias utilizadas. O facto de no artigo desenvolvido em [13], todos os *posts* terem sido colocados numa das sete categorias usadas pelos autores, permitiu-lhes a construção de um modelo estatístico com parâmetros que permitiram a avaliação do ganho em cada mensagem, o que significa que os autores possibilitaram a realização de comparação do número de *posts* em cada categoria e a resposta que obtiveram através de *likes*, *shares* e *comments*. No entanto, no estudo [13] assenta numa metodologia que faz uso de um modelo editorial prescritivo, ou seja, o modelo editorial é definido à priori e todas as mensagens publicadas em social media são classificadas de acordo com uma das sete categorias desse modelo. Embora o modelo editorial proposto possa ser diversificado em termos de áreas editoriais, estas podem não ser suficientes nem adequadas.

Segundo o estudo [3], *Chauhan e Pilai* usam um método rigoroso e uma análise sofisticada de conteúdo estatístico, mas a recolha de dados e a categorização são conduzidas, exclusivamente, por humanos, o que não é uma solução viável nem económica para organizações que procuram monitorização nas redes sociais. Por outro lado, *Lai e To*[10] sugerem uma teoria fundamentada, baseada numa análise léxica assistida por computador com métodos estatísticos e gráficos, para identificar as dimensões dos tópicos, minimizando os erros humanos, bem como os vícios de codificação e categorização. Neste âmbito, a investigação consiste na identificação de frequência das palavras-chaves mediante um *software* lexical e na aplicação de análise fatorial exploratória para agrupá-los em vários fatores. A principal limitação das técnicas de *text mining* apresentada na metodologia de *Lai e To* é que, quando consideradas isoladas/não integradas noutros processos/tecnologias mais robustos, os processos não são consistentes, desse modo os autores propõe essa integração e manipulação das técnicas.

2.4 Problemas esperados

A classificação de texto pode levantar problemas, dado que palavras ou expressões muito similares podem levar a inconsistências ou interpretações erradas, podendo originar resultados diferentes dos que inicialmente se podem esperar.

O objetivo da classificação de texto é que, através da análise de *posts*, o sistema seja capaz de conseguir encontrar, no mínimo, três categorias a que pertencem as informações recolhidas do *Facebook*. No entanto, a intenção não é que se encontrem muitas categorias com poucos elementos cada, porque isso indicaria a presença de ruído ou *outliers*, ou seja, a informação não seria plausível. Um vez resolvidas todas as possíveis inconsistências de classificação de texto, surge o fato de estar a ser implementado um classificador que não seja treinado nem tenha ajuda humana. Com este processo, é possível reter informação mais importante e útil para o estudo de forma autónoma, visto que, a partir do classificador, se tornaria possível obter grandes quantidades de informação num espaço curto de tempo, o que não acontece se for realizada com ajuda manual.

Capítulo 3

Análise exploratória

Este capítulo visa relatar detalhadamente, todo o processo de acesso aos dados, o modo de recolha de informação, a forma como esta é trabalhada e utilizada para a análise posteriormente desenvolvida. Neste capítulo, é ainda apresentado um estudo exploratório inicial acerca do conjunto de dados (*dataset*) recolhido.

3.1 Recolha e tratamento dos dados

3.1.1 Configuração de *API* do *Facebook*

Para a recolha de dados que se encontravam no *Facebook* foi necessário a implementação de algumas etapas prévias à sua realização, sendo a criação e configuração de uma *Facebook API* (figura 3.1) de carácter obrigatório, uma vez que permitiu a comunicação do programa desenvolvido com a rede social.

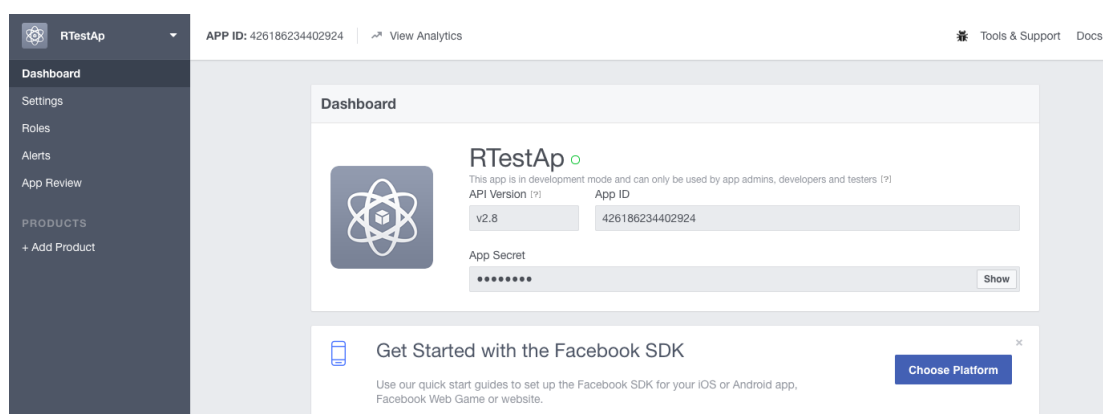


Figura 3.1: Criação da *APP* para o sistema

Após a criação da aplicação na *API do Facebook* verificou-se a necessidade de consultar as permissões da mesma, de forma a verificar se a aplicação continha um perfil público, para

assim possuir as permissões por defeito. Para isso, primeiramente selecionou-se o separador *Tools and Support*, seguidamente *Graph API Explorer* e por fim *Get Token* que permite um acesso temporário e seguro às *APIs do Facebook*. Este *Token* permite identificar o utilizador, aplicativo ou página, e é apenas válido por duas horas, uma vez que a sessão expira depois disso. Seguidamente, foi necessário verificar o separador das permissões referidas na figura 3.2. Se o perfil for público as permissões estão incluídas por defeito.

The screenshot shows the 'Select Permissions' window in the Facebook Graph API Explorer. It has a title bar with 'v2.10' and a close button. The window is divided into four sections of permissions, each with a list of checkboxes. The 'Public profile included by default' checkbox at the bottom left is checked and highlighted with a red box. The 'Get Access Token' button at the bottom right is also highlighted with a red box. Other buttons include 'Clear' and 'Cancel'.

Section	Permissions
User Data Permissions	<input type="checkbox"/> email <input type="checkbox"/> publish_actions <input type="checkbox"/> user_about_me <input type="checkbox"/> user_birthday <input type="checkbox"/> user_education_history <input type="checkbox"/> user_friends <input type="checkbox"/> user_games_activity <input type="checkbox"/> user_hometown <input type="checkbox"/> user_likes <input type="checkbox"/> user_location <input type="checkbox"/> user_photos <input type="checkbox"/> user_posts <input type="checkbox"/> user_relationship_details <input type="checkbox"/> user_relationships <input type="checkbox"/> user_religion_politics <input type="checkbox"/> user_status <input type="checkbox"/> user_tagged_places <input type="checkbox"/> user_videos <input type="checkbox"/> user_website <input type="checkbox"/> user_work_history
Events, Groups & Pages	<input type="checkbox"/> ads_management <input type="checkbox"/> ads_read <input type="checkbox"/> business_management <input type="checkbox"/> manage_pages <input type="checkbox"/> pages_manage_cta <input type="checkbox"/> publish_pages <input type="checkbox"/> pages_messaging <input type="checkbox"/> pages_messaging_payments <input type="checkbox"/> pages_messaging_phone_number <input type="checkbox"/> pages_messaging_subscriptions <input type="checkbox"/> pages_show_list <input type="checkbox"/> read_page_mailboxes <input type="checkbox"/> rsvp_event <input type="checkbox"/> user_events <input type="checkbox"/> user_managed_groups <input type="checkbox"/> pages_manage_instant_articles <input type="checkbox"/> user_actions.books <input type="checkbox"/> user_actions.fitness <input type="checkbox"/> user_actions.music <input type="checkbox"/> user_actions.news <input type="checkbox"/> user_actions.video
Other	<input type="checkbox"/> instagram_basic <input type="checkbox"/> instagram_manage_comments <input type="checkbox"/> instagram_manage_insights <input type="checkbox"/> read_audience_network_insights <input type="checkbox"/> read_custom_friendlists <input type="checkbox"/> read_insights

Figura 3.2: Permissões da API

Depois de ter em conta as permissões, tivemos o acesso ao *Token* que permitiu a conexão entre o programa desenvolvido em R para extrair informação e a *API do Facebook*

3.1.2 Extração dos dados

Terminadas as etapas de criação e configuração da *API do Facebook* descritas em cima, passou-se à extração dos dados. Neste processo, contou-se com o recurso ao *package Rfacebook*, que permitiu a utilização de funções, uma vez que é uma coleção de funções R, códigos e dados. Depois de instalado o *package*, utilizou-se o *Token* para permitir a conexão entre o programa desenvolvido e a *App do Facebook*.

Posteriormente, utilizaram-se os IDs das instituições para aceder às páginas oficiais. Foram recolhidos 1100 registos por instituição. A informação recolhida foi guardada num *dataset*, sendo esta obtida em 45 minutos, e totalizando 11008 registos. Nestes registos recolheram-se os campos ID do *post*, contagem de likes, ID da universidade, nome da universidade, mensagem publicada, data e hora da publicação, tipos de ligação publicada, link utilizado, *story*, número de *comments*,

shares, love, haha, wow, sad e angry.

3.2 Tratamento de dados

Uma vez que todos os dados foram recolhidos e inseridos num *dataset*, isto é, num conjunto de dados que corresponde ao conteúdo de uma única tabela de base de dados, em que cada coluna da tabela representa uma variável particular e cada linha corresponde a um dado do conjunto de dados em questão, sucedeu-se o tratamentos dos mesmos. Inicialmente, foi criada uma variável que apenas continha a data de publicação da mensagem. Tendo esta informação, uma vez que tencionava estudar as publicações realizadas durante um ano letivo, as mensagens teriam de ser publicadas entre o período de 1 de setembro de 2015 a 31 de agosto de 2016. Tendo isto em conta, todas as publicações recolhidas que não faziam parte deste período foram eliminadas, ficando apenas 5142 registos no *dataset* para serem analisadas.

Uma vez que os dados contidos no *dataset* são os que estão dentro do período pretendido, foi adicionada uma nova variável à tabela que indicava o número de palavras que continha cada mensagem publicada na rede social. Posteriormente, foram removidos todos os *emojis* e a mensagem foi transformada num *VCorpus*, usado para criar objetos da classe *Corpus* que contém um conjunto de dados baseado em texto, para remover a pontuação, os números, as *stopwords* e os espaços em branco extra, ou seja múltiplos caracteres de espaços em branco são colapsados num único. Todas as letras maiúsculas foram convertidas em minúsculas e por fim reduziram-se as palavras ao mesmo *stem*. Por fim, o *VCorpus* foi convertido em *dataframe*, que é um suporte usado para armazenar tabelas sendo que cada coluna da tabela representa uma variável e a linha corresponde a um dado da tabela. Este *dataframe* foi adicionado numa variável ao *dataset* inicial.

Tendo em conta o *dataset* com as novas variáveis inseridas, foram obtidos os dados numéricos representados na tabela 3.1.

	likes_count	comments_count	shares_count	love_count	haha_count	wow_count	sad_count	angry_count	word_count
Min.	1,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,0
1st Qu.	178,2	5,0	15,0	0,0	0,0	0,0	0,0	0,0	17,0
Median	428,0	12,0	43,0	3,0	0,0	0,0	0,0	0,0	26,0
Mean	1057,4	31,33	140,9	19,61	1,246	6,616	4,425	0,5592	29,7
3rd Qu	1140,8	32,00	110,0	14,00	0,0	4,0	0,0	0,0	37,0
Max	42674,0	5323,0	85935,0	1612,0	510,0	1099,0	6683,0	381,0	322,0
NA's	-	-	-	1	1	1	1	1	96

Tabela 3.1: Síntese dados numéricos

Como se verifica na tabela 3.1, existem valores não disponibilizados (*NAs*), os quais foram convertidos a zero, com o propósito de não encontrar limitações na manipulação dos dados. Os *ID's* de cada universidade converteram-se para o tipo numérico e colocado numa nova variável no *dataset*, para facilitar a manipulação dos dados.

3.3 Análise dos dados recolhidos

Com os dados recolhidos através do *Facebook App* aplicaram-se métodos com as variáveis numéricas de modo a que se conseguisse perceber o comportamento das universidades ao longo do ano letivo. Posteriormente, realizaram-se somatórios para constatar o comportamentos das instituições a nível de número de publicações, as interações que elas geraram junto do público-alvo e a preferência de publicação em relação aos dias da semana e ao mês do ano. Uma vez realizados estes processo, procedeu-se à análise individual do comportamento de cada universidade. Neste ponto os dados que estão a ser trabalhados são separados por instituição, usando um *dataset* por entidade. Esses *datasets* são analisados tendo em conta o calendário escolar¹ de cada instituição. Finalmente, utilizou-se correlação de *Pearson* para encontrar eventuais relações fortes entre as diversas variáveis numéricas do *dataset*.

No que se refere ao número de publicações realizadas pelas instituições é possível verificar que, no período de 1 de setembro de 2015 a 31 de agosto de 2016, a instituição com mais publicações é a Columbia University com 740 *posts*, seguindo-se a UC Berkely com 681 *posts*, Massachusetts Institute of Technology (MIT) (com 621 *posts*) e Yale University (com 598 *posts*) são as duas instituições com o número de publicações mais aproximadas. Seguidamente, encontra-se a universidade de University of Cambridge (com 563 *posts*) e Harvard University (com 529 *posts*). É possível verificar números muito próximos de publicações nas três universidades seguintes, Princeton University (com 439 *posts*), University of Oxford (com 377 publicações) e University of Chicago (apresentando um total de 359 *posts*). A instituição que se destaca pela negativa, entre as dez estudadas, é Stanford University com um somatório de 235 *posts*, como ilustrado na figura 3.3.

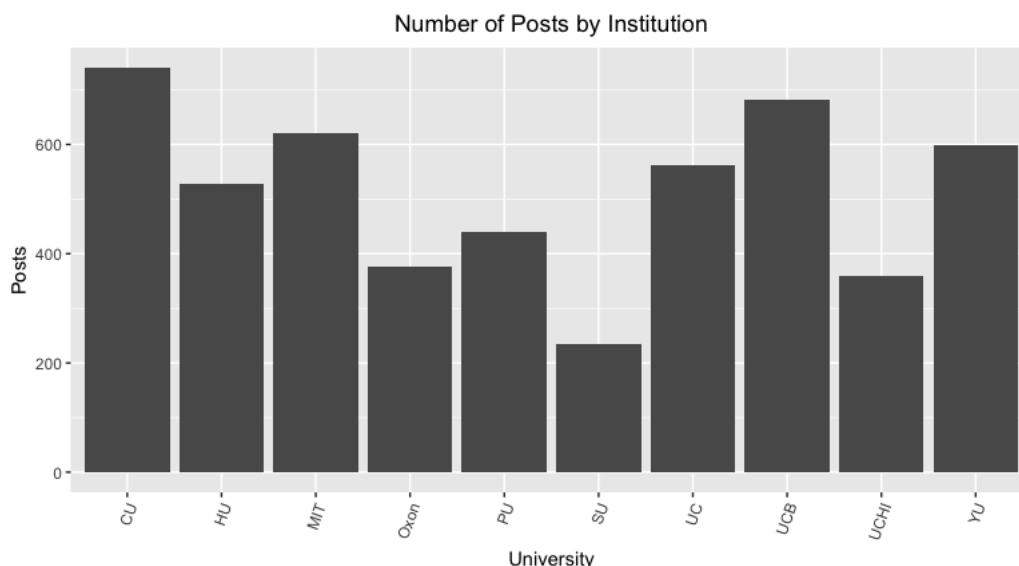


Figura 3.3: Número de *posts* por Instituição

¹disponíveis em anexos

Apesar de a Columbia University ser a universidade com maior atividade no *Facebook*, é uma das instituições que apresenta valores mais baixos de interações geradas com os fãs (soma total de *comments*, *likes*, *shares* e *reactions*), com um total de 181684, juntamente com Stanford University (com 179029 interações). A Harvard University lidera na capacidade de gerar maior número de interações no *Facebook*, apresentando valores distintamente mais elevados do que as restantes universidades, tal como é possível observar na figura 3.4 (mais de 2 milhões de interações). As instituições com valores mais aproximadas registam cerca de metade deste valor e são a Universidade de University of Oxford (1258035 interações) e a UC Berkeley (com 853327). University of Cambridge aparece posteriormente com 655586 interações, seguida por Massachusetts Institute of Technology (MIT) com 499979 interações. Por fim, com menor número de interações temos as universidades de Yale University (326496 interações), Princeton University (262690 interações) e University of Chicago (com um total de 208878 interações). Note-se que, regra geral, a quantidade de interações são reduzidas.

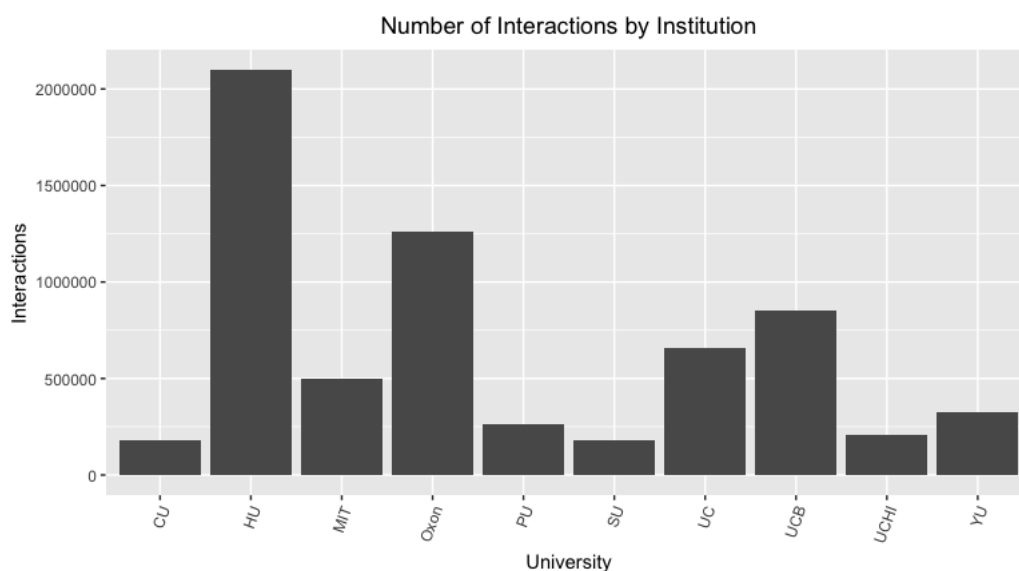


Figura 3.4: Número de interações por Instituição

No que diz respeito à contagem de palavras que constituem a mensagem publicada, Columbia University e Yale University são as instituições que mais palavras usam nas publicações, 24084 palavras e 22746 respetivamente. Posteriormente apresentam-se as universidades Harvard University (com 10664 palavras), University of Chicago (usando 11213 palavras) e Massachusetts Institute of Technology (MIT) (com 12816 palavras), sucedendo a University of Cambridge que utiliza 15321 palavras, University of Oxford com 17721 palavras e UC Berkeley com um total de 18665 palavras. Por fim, a Stanford University que usa menos palavras nas suas publicações (7870 palavras) seguindo-se de Princeton University (8756 palavras), como ilustrado na tabela 3.2.

Em relação ao período de publicações, nota-se uma grande preferência na publicação de conteúdos durante os dias úteis da semana, tendo a segunda-feira uma frequência mais baixa comparada com os restantes dias. A quarta-feira é o dia com mais publicações por parte das

Universidades	Número de palavras
CU	24084
YU	22746
UCB	18665
Oxon	17721
UC	15321
MIT	12816
UCHI	11213
HU	10664
PU	8756
SU	7870

Tabela 3.2: Número de palavras por Instituição

instituições, com um total de 888 *posts*, seguindo-se a terça-feira e a quinta, que registam o mesmo número, 866 *posts*. A sexta-feira é o segundo dia útil que conta com o menor número de publicações registadas, contando com 893 *posts* realizados. Como se verifica na tabela 3.3, aos sábado realizaram-se 465 *posts*, seguindo-se o domingo com 395 *posts*, o dia que menos publicações apresenta.

Dias da semana	Número de publicações
Quarta-feira	888
Quinta-feira	866
Terça-feira	866
Sexta-feira	863
Segunda-feira	809
Sábado	465
Domingo	385

Tabela 3.3: Número de publicações por dia

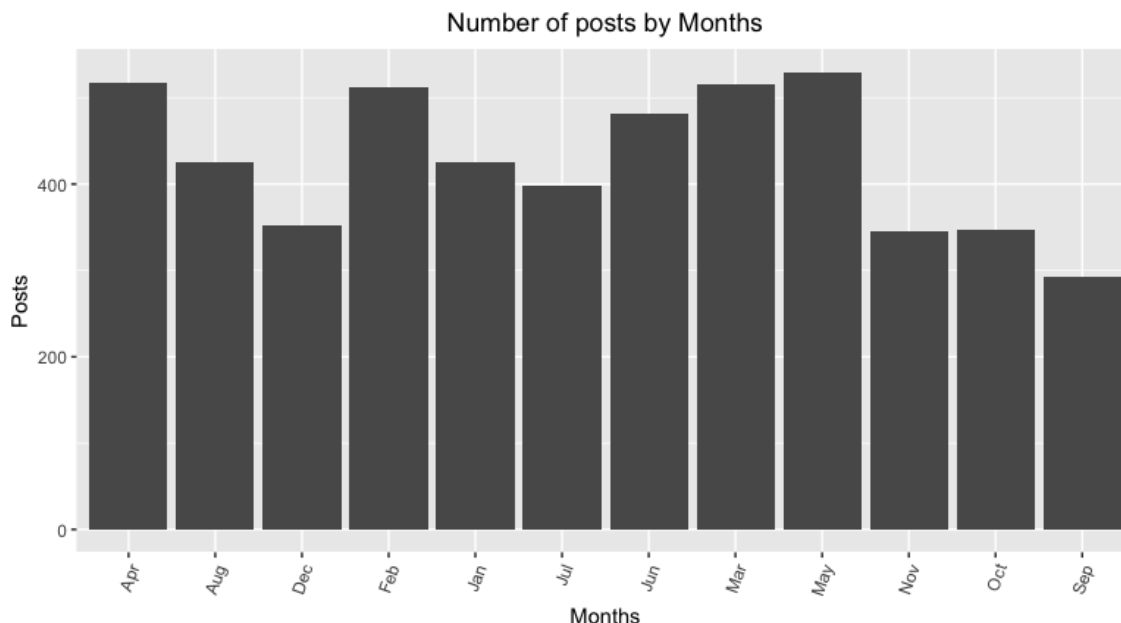
No mesmo estudo realizado para os meses do ano, foram detetados quatro meses em que o número de publicações é muito aproximado: o mês de maio lidera a contagem com 530 *posts*. Abril, março e fevereiro com 517, 515 e 512, respetivamente, sucedendo-se os meses de junho, agosto, janeiro e julho. O mês de dezembro (com 352 *posts*), outubro (com um total de 347 *posts*) e novembro (num total de 345 *posts*), como ilustrado em 3.5. É importante salientar, que o mês de setembro regista menos *posts* divulgados (com 293 *posts*).

Como mencionado em cima, realizou-se um estudo exploratório inicial com o propósito de estudar cada universidade e para se poder perceber o seu comportamento individual ao longo do ano letivo, com recurso ao calendário escolar ².

A Harvard University, em média, no período de férias realiza um a dois *posts* por dia, mas, por vezes registam três publicações. O máximo de *posts* diários foi de sete publicações registadas no dia 26 de maio de 2016, que corresponde ao dia da formatura, seguindo-se o dia anterior, a véspera do dia da formatura, com 5 *posts*.

A Harvard University apresentou uma média de 1,511 publicações diárias às quais se registaram uma média de 5223 interações. Os *posts* com menos *reactions* não pertencem unicamente às interrupções letivas, uma vez que durante alguns dias, durante o período letivo, se registaram poucas interações, nomeadamente no dia 28 de fevereiro de 2016, onde foram perceptíveis 317

²disponíveis nos anexos

Figura 3.5: Número de *posts* por meses do ano

reactions.

A Stanford University torna-se difícil de analisar, uma vez que a instituição em causa tem uma relação escassa com as redes sociais. Pelo que foi verificado, os fins-de-semana são muito propícios à não existência de publicações, tal como as pausas letivas que também possuem poucas publicações. A Stanford University registou uma média de 1,243 publicações diárias às quais se verificaram uma média de 796,1 *reactions*. Através do estudo da presença desta universidade nas redes sociais, constatámos que no dia 6 de dezembro de 2015, referente à época de exames, se registaram 5 *posts*, sendo assim o dia com mais publicações, às quais se registaram 6658 *reactions*, tornando-se assim o segundo dia com maior número de *reactions*. O dia em que se registou o maior número de *reactions* foi o dia 2 de janeiro de 2016, com 7501 interações sendo referentes à abertura da residência de graduação para o trimestre de inverno.

A Massachusetts Institute of Technology publicou em média 2,023 *posts* por dia às quais se registaram 1347 *reactions*. No entanto existe um período de 1 de setembro de 2015 a 13 de outubro de 2015 onde não existe qualquer tipo de atividade na página. Por isso, é possível concluir que a instituição não utilizava esta rede social antes de 13 de outubro de 2015. É possível afirmar que a universidade em estudo é muito participativa na rede social *Facebook*, apresentando cinco dias com cinco *posts*, sendo um deles o dia 1 de junho de 2016, correspondente à véspera de cerimónia de doutoramento. É ainda importante frisar que os dias com menos *reactions* dizem respeito aos dias com menos *posts*.

A University of Cambridge, mesmo em interrupções letivas, utiliza o *Facebook* como forma de interação com o público. Os dias em que não existe nenhum *post* ou que o seu número é bastante reduzido são os sábados e domingos, ou seja os fins-de-semana. Por vezes, também as sextas e segundas são propícias à não existência de publicações. É possível verificar que University of

Cambridge apresentou uma média de 2,298 publicações às quais se registaram 2315 *reactions*. Os dias com mais *posts* foram os dias 3 de maio e 26 de maio de 2016, com seis *posts* cada. Em relação ao maior número de *reactions* destacaram-se os dias 16 de dezembro de 2015 com 11395 *reactions*. É de frisar que os dias com maior número de *posts* não dizem respeito aos dias com maior número de *reactions*, uma vez que no dia 16 dezembro de 2015 apenas foram publicados 3 *posts* e no dia 20 de maio de 2016 apenas se registaram 4 publicações.

Ao contrário das outras instituições, a University of Oxford tem uma atividade regular no *Facebook*, inclusive nos meses de julho e agosto, onde houve atividade diária. O mês de abril, é o que possui menos publicações. Contudo, apenas não publicaram durante três dias, talvez porque nesse mês é a mudança de trimestre. É possível verificar também que no único dia com três *posts*, dois deles são repetidos e foram publicados com diferença de um minuto. O *post* do dia 23 de abril só tem *likes*, não sendo registadas outras *reactions*. A universidade de Oxford registou uma média de 1,071 publicações, obtendo uma média de 2798 *reactions* por dia. Note-se que nesta universidade, os dias com mais *reactions* são registados em dias com menos *posts*.

A instituição Columbia University começou o ano letivo em agosto, mas até 14 de setembro, não manteve qualquer atividade no *Facebook*, temporada em que se deu início à mudança de programa. O 18 de maio, dia com maior número de *reactions* e também um dos dias com maior número de *posts*, diz respeito ao regresso de todos os estudantes do ensino profissional e à formatura. A média de *posts* diárias desta universidade é 2,587 obtendo uma média de 546,1 *reactions*. Esta instituição opta por não publicar com tanta frequência aos fins-de-semana.

Na University of California, Berkeley, são raros os dias em que não existe uma publicação. A partir do momento em que começam as férias académicas e administrativas, contabiliza-se um decréscimo de publicações, ou seja, durante os meses de junho, julho e agosto. UC Berkeley apresentou uma média diária de 1,94 publicações às quais se registaram 2163 *reactions*. Na página desta universidade observam-se dezasseis dias com quatro *posts*, destacando-se os dias 9 de maio referente ao fim dos exames de primavera e o dia 13 de maio de 2016, correspondente ao fim do semestre de primavera. Este último dia, 13 de maio de 2016, foi o segundo dia com mais *reactions*, registando 12594, existindo apenas o dia 8 de fevereiro de 2016 com 22080.

A University of Chicago aos fins-de-semana, não tem tantos *posts*, sobretudo aos domingos. Em interrupções letivas mostra um comportamento igual às outras alturas do ano. O dia com mais *posts* (11 de junho) corresponde ao dia em que termina o quadrimestre de primavera e o dia da cerimónia de "convocação". Em relação ao número de *reactions*, o dia com maior número de *reactions*, 4 de janeiro, é o dia que se inicializa o quadrimestre de inverno. A 11 de fevereiro, segundo dia com mais *reactions* é a véspera da pausa escolar. A média diária desta universidade foi de 1,234 publicações às quais se registaram 600,6 *reactions*.

A atividade nas redes sociais da Princeton University, iniciou-se a 20 de janeiro de 2016. A partir desta altura manteve um papel bastante ativo, passando a não publicar apenas aos fins-de-semana. O mês de agosto foi o mês em que as publicações diminuíram, visto que foi a única altura do ano em que não existiram publicações durante a semana. O dia com mais

posts, 5 de fevereiro de 2016, diz respeito ao início do trimestre de primavera, como é referido no calendário escolar ³, sendo também um dos dias com mais *reactions*. O dia 16 de março corresponde ao dia com mais *reactions*, 5788, no que diz respeito às férias de primavera. Esta universidade apresentou uma média de 2,152 publicações diárias e uma média de 1186,2 *reactions*.

A Yale University, só começou a atividade a 29 de dezembro de 2015. Desde aí registou-se apenas um dia sem publicações. A universidade de Yale registou uma média de 2,431 *posts* diários obtendo uma média de 1169,8 *reactions*.

Por ultimo, desenvolve-se um *script* que permite calcular a correlação de *Pearson* entre as variáveis empregues no estudo. Note-se que na correlação de *Pearson* quando os valores de ambas as variáveis aumentam, obtemos uma correlação perto do valor 1, sendo assim linear perfeita ou diretamente proporcional. Por outro lado, quando uma variável aumenta e a outra diminui obtemos uma correlação de valor -1, sendo também linear perfeita mas inversa. Com a análise do *dataset*, apenas se encontram correlações positivas, realizando-se agrupamentos das diferentes *reactions*, para ser possível entender melhor as variáveis e o conjunto de dados em estudo. É definido um valor mínimo de correlação, o que possibilita verificar quais as variáveis que, de uma forma ou de outra, se correlacionam. Com a criação de agrupamentos, como é o caso do agrupamento do número de todas as *reactions*, incluindo os *likes*, verifica-se que, à medida que este agrupamento cresce, o número de *comments* também cresce. O mesmo se passa com o agrupamento de *reactions* positivas (o número de *likes*, de *loves*, de *haha* e de *wow*), de tal modo que, se este agrupamento cresce, os número de *comments* também crescem. Verifica-se então que o número de *likes* aumenta com número de *comments*, do mesmo modo que o número de *shares* aumenta com o aumento do número de *comments*.

É criado um agrupamento, que inclui todos as *reactions* negativas, incluindo os *likes*, os *haha* e os *wow*, visto que estas *reactions* podem ser usadas para expressar os ambos polos, negativo e positivo. Este agrupamento cresce à medida que o número de *comments* aumenta. Outro agrupamento criado para o estudo das correlações é a junção das *reactions* *wow* e de *haha*, que, quando crescem, cresce a *reaction love*, o mesmo se passa com a *reaction wow*, que aumenta com o aumento de *reaction love*, e a *reaction angry*, que também aumenta com o aumento da *reaction angry*.

Nos restantes agrupamentos gerados, um passa por incluir as *reactions* *haha* e *wow* e outro inclui as *reactions* *wow* e *angry*, em que ambas aumentam de acordo com o aumento da *reaction love*. Também é possível observar que os valores deste último agrupamento (*wow* e *angry*) têm tendência a subir em conformidade com o agrupamento de *reactions love* e *haha*.

³disponível em anexo

3.4 Conclusão

Com isto, é possível concluir que os agrupamentos gerados para a correlação de *Pearson* não nos transmite muita informação, posto que uma das variáveis dos agrupamentos já se relaciona sozinha com a variável em causa, não sendo encontrada nenhuma variável que dependa diretamente de outra, positivamente, visto que o valor mais alto encontrado foi de 0.8 e não foi encontrado nenhuma correlação com o valor máximo, 1. Em relação às atividades das instituições, em geral todas elas possuem uma atividade ativa com o *Facebook*, à exceção de Stanford University que apresenta uma atividade bastante escassa. Normalmente, os seguidores das instituições são extremamente participativos.

Capítulo 4

Análise de comportamento e desempenho

Nesta capítulo focamos-nos na exploração dos métodos de verificação de comportamento e desempenho por universidade. As instituições são examinadas a nível de publicações realizadas e de atividade na rede social existente ao longo do tempo entre os seguidores e as páginas das entidades. Por fim, é analisado o desempenho das universidades por trimestre, no qual se estuda a performance das instituições e se relacionam as publicações e as interações por parte da comunidade académica.

4.1 Evolução do comportamento ao longo do tempo

A informação recolhida através do *Facebook App* proporciona a realização de uma discretização ou suavização de dados ruidosos, ou seja, agrupamento dos valores relacionados em repartições, para reduzir o número de valores distintos, fortalecendo assim a relação entre os atributos. Este procedimento é praticado para o número de publicações realizadas diariamente pelas instituições envolvidas no estudo e para o número de interações obtidas por elas (soma de *likes*, *comments* e *reactions*).

Na suavização dos dados, os valores são distribuídos por grupos, a que se dá o nome de *bins*, que mantêm a representatividade dos elementos. Em cada grupo, é aplicado um processo de escolha de uma medida para serem ajustados, tal como a média aritmética, mediana ou um valor limite. Os valores originais são substituídos pelas medidas calculadas em cada grupo [8].

Inicialmente, os dados são agrupados pela quantidade de *posts* diários e é calculada a sua média. Esses agrupamentos são produzidos para três dias, sete dias e um agrupamento para distinguir a quantidade de publicações praticadas durante os dias úteis da semana e o fim-de-semana, de modo a encontrar diferenças entre essas alturas. Este estudo foi igualmente realizado para a interação entre os seguidores e as páginas oficiais das instituições, de forma a estudar o desempenho das instituições. Escolhe-se a melhor visualização de agrupamentos para reportar

uma melhor evolução e transmitir mais informação de cada entidade, tendo como auxiliar o calendário escolar e a página no *Facebook* de cada universidade.

4.1.1 Atividade no *Facebook* ao longo do tempo

Através da análise da figura 4.1, constatamos que Harvard University mantém alguma consistência na atividade, tendo uma média diária de 2 *posts*, trata-se, portanto de uma instituição que publica com regularidade, apresentando um comportamento consistente ao longo de todo o período de análise. Note-se que há uma ligeira tendência para diminuir os *posts* que são feitos em dias úteis. Regista-se uma maior intensidade de publicações durante as primeiras treze semanas do ano letivo, seguidas de um decréscimo até por volta da trigésima semana. A partir dessa altura e até à trigésima nona semana, a intensidade de publicações volta a aumentar, verificando-se aqui o pico mais alto do número de publicações, pelo que se assume que as publicações sejam relacionadas com este tema. Após esta semana, a intensidade de publicações volta a diminuir. No início das férias de primavera, que começam num fim-de-semana, a instituição não realizou qualquer tipo de publicação. Contrariamente ao que acontece com a frequência de publicações em dias úteis, a instituição demonstra uma tendência para aumentar o número de publicações ao fim de semana. Durante as primeiras treze semanas, em que há um comportamento mais consistente de publicações em dias úteis, o número de *posts* ao fim-de-semana é mais reduzido. No entanto, após estas treze semanas, à medida que vão variando os períodos de intensidade de publicações em dias úteis, a tendência aponta para um aumento (sensivelmente o dobro) do número médio de publicações feitas ao fim-de-semana, mantendo-se este comportamento relativamente consistente até ao final do ano letivo.

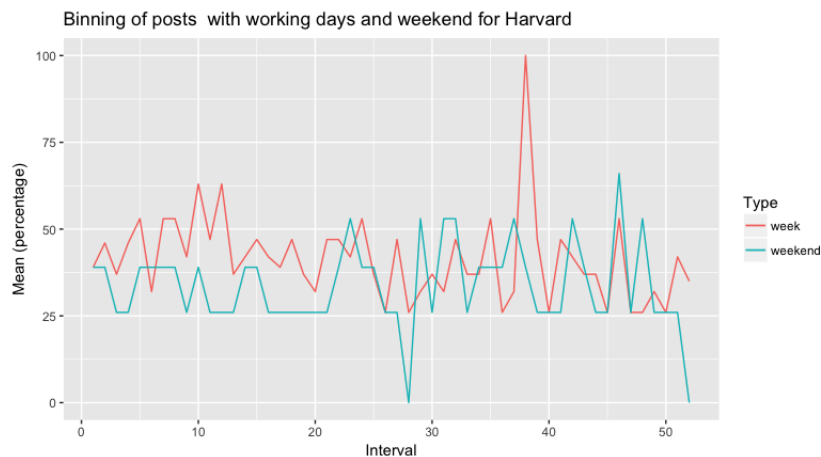


Figura 4.1: Agrupamento de *posts* para *Harvard*

Na figura 4.2 foi adotada uma visualização com agrupamento de três dias. Através da sua análise, constatamos que Stanford University tem pouca tendência para a publicação, apresentando uma média de 1,2 *posts* diários. Até ao agrupamento dos dias 1,2 e 3 de dezembro de 2015 (décimo sétimo ponto) verificamos uma atividade constante, havendo um aumento de *posts* nos dias 4, 6 e 8 de dezembro de 2015 (ponto dezoito), sendo que estes dias correspondem

aos últimos dias de aulas, últimos dias para enviar a tese, prazo tardio para conferência do grau de trimestre de outono, exames na faculdade de direito e competições de futebol. O pico de maior número de *posts* acontece no ponto vinte e quatro, correspondente ao conjunto dos dias 1, 2 e 7 de janeiro de 2016, sendo estes o primeiro dia do ano civil, o dia da abertura da residência para o trimestre de inverno e o dia de conferência de graus e final do trimestre de outono, respetivamente. Após este pico, há uma tendência para a ligeira diminuição do número de *posts*, acentuando-se o ponto vigésimo nono, referente ao conjunto dos dias 4, 5 e 7 de fevereiro de 2016, o ponto trigésimo segundo, que corresponde aos dias 24, 25 e 26 de fevereiro e ponto quadragésimo terceiro, o que diz respeito aos dias 24, 25 e 26 de abril. Após estes pontos, há uma diminuição de publicações até que se encontra um comportamento consistente a partir do ponto quinquagésimo oitavo (correspondente ao conjunto dos dias 20, 21 e 25 de julho) até ao final do ano letivo, verificando-se uma média de apenas 37% de *posts*. Note-se ainda que as publicações de Stanford chegam a atingir variações de 60%.

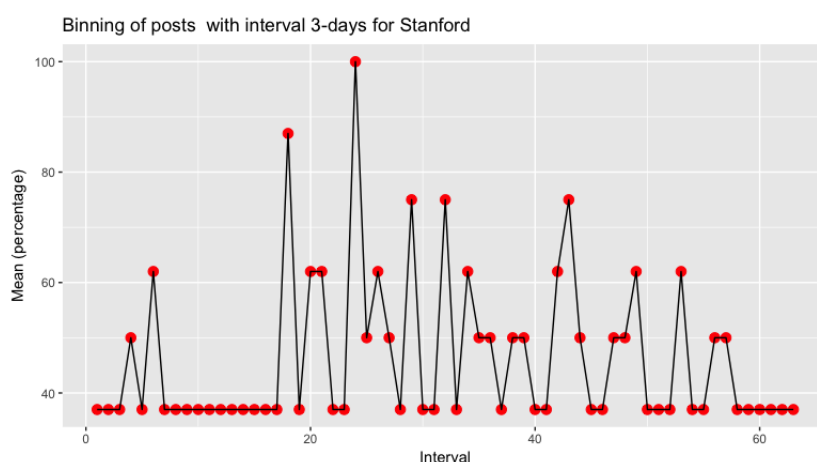


Figura 4.2: Agrupamento de *posts* para *Stanford*

Na figura 4.3 é visível um *binning* de sete dias referente às publicações de MIT. Este revela um comportamento aleatório, apresentando algumas variações significativas, como se pode verificar na figura 4.3. Até ao ponto 10, vão-se registando aumentos e diminuições de *posts* sem grande relevo, até chegarmos ao ponto onze, referente aos dias 28, 29, 30 e 31 de dezembro e 1, 2 e 3 de janeiro, onde se regista uma diminuição abrupta de *posts*, chegando aos 41%. Após esta diminuição, vai-se registando um aumento ligeiro de publicações, à exceção do ponto vinte e um. Em seguida, verificamos um aumento abrupto, seguido de uma brusca diminuição de *posts*, criando um efeito de arco de volta perfeita. Este período inicia-se no ponto vigésimo sexto, (com início em 14 de abril), referente ao dia do Patriota, à Maratona de *Boston* e ao início das férias e termina no ponto trinta e sete (com início em 3 de julho) correspondente à entrega das teses e início do período de exames. Note-se que o pico de maior número de *posts* diários é atingido nos pontos trinta e um e trinta e dois, alcançando os 100%. Estes pontos estão relacionados com as férias escolares e com a formatura.

A figura 4.4 apresenta a atividade da página de University of Cambridge através do uso de um *binning* de dias úteis e fins de semana. É importante frisar que University of Cambridge

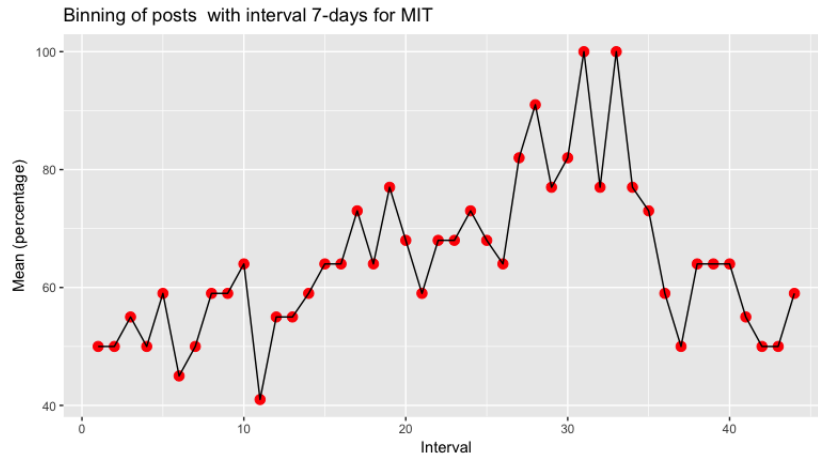


Figura 4.3: Agrupamento de *posts* para MIT

apresenta uma atividade bastante irregular, apresentando uma média diária de 2 *posts* por dia. Ao longo da análise do gráfico, é notável a existência de uma zona distinta na semana treze, que diz respeito aos dias 28, 29 de abril e 3 de maio de 2016, sendo estes a segunda feira de Páscoa e o início do novo período. Este período corresponde ao pico de maior número de publicações (100%). Após este pico, há uma brusca descida, até à semana catorze e novamente aumentos e diminuições abruptas. Salienta-se ainda a vigésima sétima semana do ano civil, correspondente aos dias 19, 20, 21 e 22 de julho, como sendo a semana com menor número de publicações (35%). Note-se que a atividade desta instituição ao fim-de-semana é muito escassa, variando sempre a quantidade de *posts* entre 0 e 1, sendo na maior parte das vezes 0.

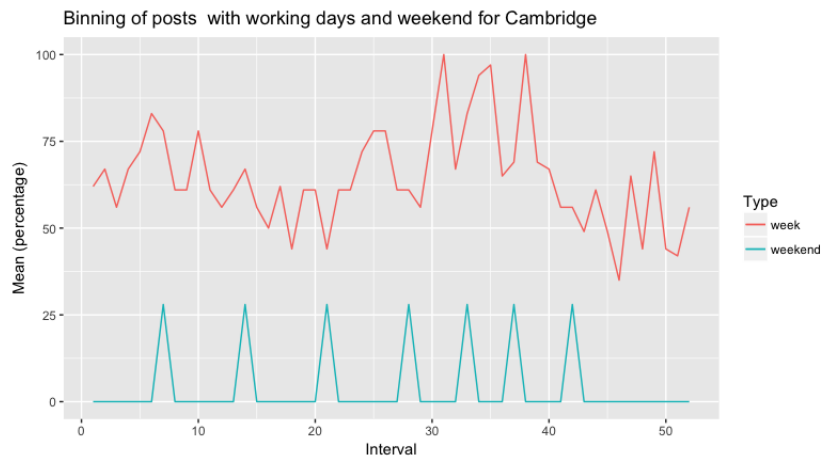


Figura 4.4: Agrupamento de *posts* para Cambridge

Na figura 4.5 é observável o registo das publicações de University of Oxford através do uso de um *binning* para três dias. Esta instituição faz publicações diariamente, mostrando uma certa obrigação em manter ativa a sua presença na rede social, obtendo uma média de um *post* por dia. No gráfico desta instituição há três períodos que se destacam, sendo o principal o ponto noventa e um, atingindo o pico de maior número de *posts* (100%). Este período corresponde aos dias 11, 12 e 13 de junho, havendo 2 *posts* repetidos. Os outros dois períodos que se destacam

são os pontos oito e noventa e cinco, ambos com 85%. O ponto oito diz respeito aos dias 23, 24 e 25 de setembro de 2015, últimos dias de férias enquanto que o ponto noventa e cinco corresponde aos dias 23, 24 e 25 de junho, últimos dias de aulas. Os restantes *posts* variam apenas entre 67% e 50%.

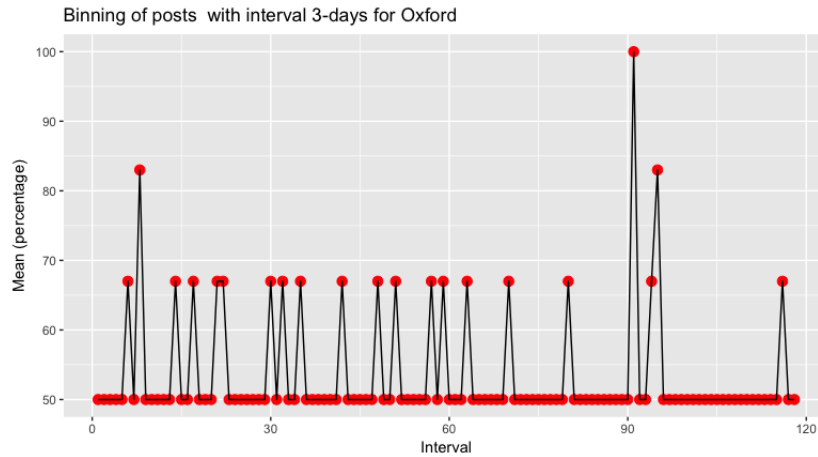


Figura 4.5: Agrupamento de *posts* para *Oxford*

No gráfico de *binning* de sete dias da figura 4.6 estão refletidas as publicações de Columbia University ao longo de um ano letivo. O primeiro momento de destaque é o ponto cinco, correspondente ao período entre 20 e 26 de outubro de 2015, onde se verifica 96% de publicações. Neste período está incluído a época de conferência de graus de outubro. É visível o pico de maior número de *posts* no ponto trinta, atingindo os 100%. Após este pico, é perceptível uma quebra de quase 50% nas publicações, sentida na trigésima semana que diz respeito ao período entre 27 de julho e o final dos dias considerados. Este intervalo diz respeito à fase de regresso dos estudantes do ensino profissional e registos de verão para o trimestre de outono

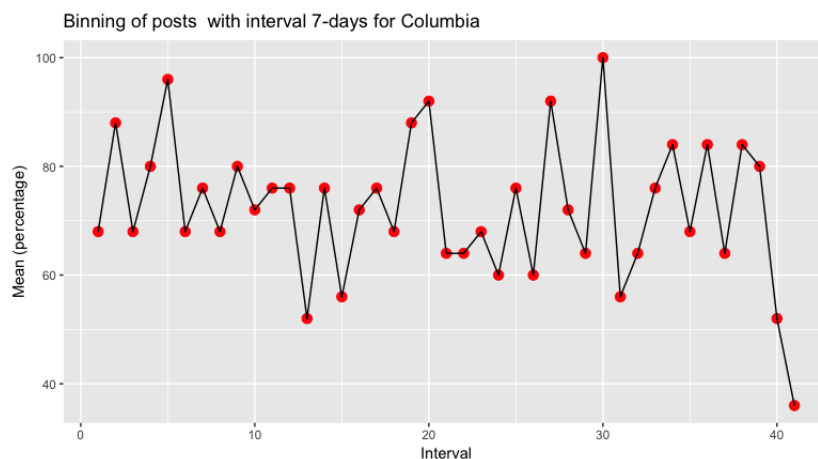


Figura 4.6: Agrupamento de *posts* para Columbia

A figura 4.7 apresenta-nos uma análise das publicações referentes à UC Berkeley utilizando um *binning* de sete dias. Analisando-a, podemos concluir que a instituição é muito ativa apresentando uma média de 2 *posts* por dia. Em relação ao número de *posts*, é possível concluir que é uma

atividade regular, visto que há aumentos e diminuições significativas. É de frisar que o ano letivo se inicia em fase de crescimento, aumentando o número de *posts*, até que aparece a primeira variação significativa, verificando-se uma diminuição brusca no ponto sete (sétima semana). Nesta semana foi o último dia de instrução, o final das aulas do trimestre de outono e a semana de leitura/ revisão/ recitação. Nesta variação encontra-se também a época de exames, seguindo-se as férias académicas e administrativas até ao início do semestre de primavera e o início de instrução. Após esta descida, regista-se novamente um aumento nos pontos oito, nove, dez e doze havendo uma quebra no ponto onze, correspondente às férias académicas e administrativas. Após a primeira semana regista-se um período de poucas publicações, entre os pontos treze e vinte e dois, voltando novamente a ter um aumento brusco no ponto vinte e três. Nota-se ainda uma grande quebra de publicações na altura do final de trimestre da primavera e o último dia de instrução, exames finais, dia de formatura e início da primeira sessão de seis semanas de cursos de verão, no ponto trinta e seis, correspondendo aos dias entre 6 de maio e 12 de maio de 2016.

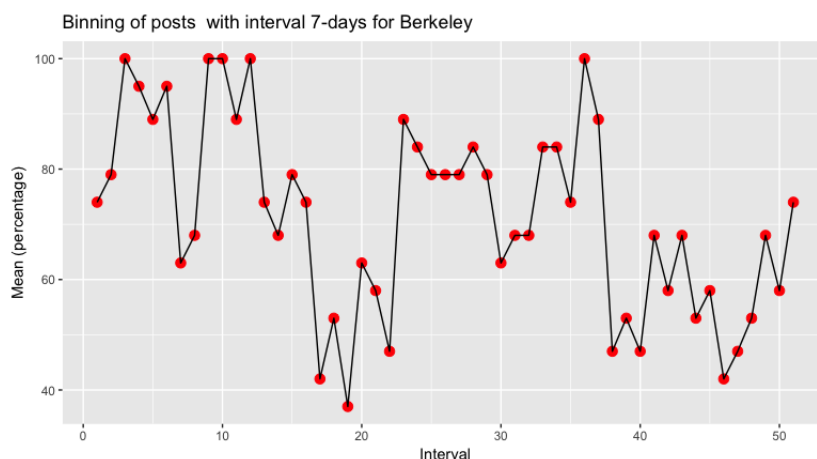


Figura 4.7: Agrupamento de *posts* para *Berkeley*

A figura 4.8 apresenta a atividade da página de University of Chicago através do uso de um *binning* de dias úteis e fins-de-semana. Esta instituição apresenta pouca tendência para publicação apresentando uma média diária de 1 *post*. Durante a semana a atividade de Chicago é regular, visto que, em média, nos dias úteis publica um *posts* por dia e ao fim-de-semana 0,6 *posts* diários. O pico de maior número de *posts* acontece na quadragésima semana, correspondente ao conjunto de dias 11 e 12 de junho de 2016 onde apresenta uma taxa de publicação de 100%, sendo nestes dias final do trimestre de primavera e ao dia de convocação. Outro fim-de-semana que se destaca neste gráfico corresponde aos dias 20 e 21 de setembro de 2015, que diz respeito ao início da semana de orientação dos novos alunos na faculdade, apresentando uma taxa de 33%. Os outros fins-de-semanas variam entre os 22% e os 0%. Durante a semana, a percentagem do número de *posts* realizados varia entre os 36% e os 22%.

Na figura 4.9, que diz respeito a Princeton University foi adotada uma visualização com agrupamento de dias úteis e fins-de-semana. Através da sua análise verifica-se uma grande tendência para publicação apresentando uma média de 2,152 *posts* diários, onde a semana apresenta uma preferência de 2 *posts* e ao fim-de-semana realiza, em média 1 *post*. Nesta

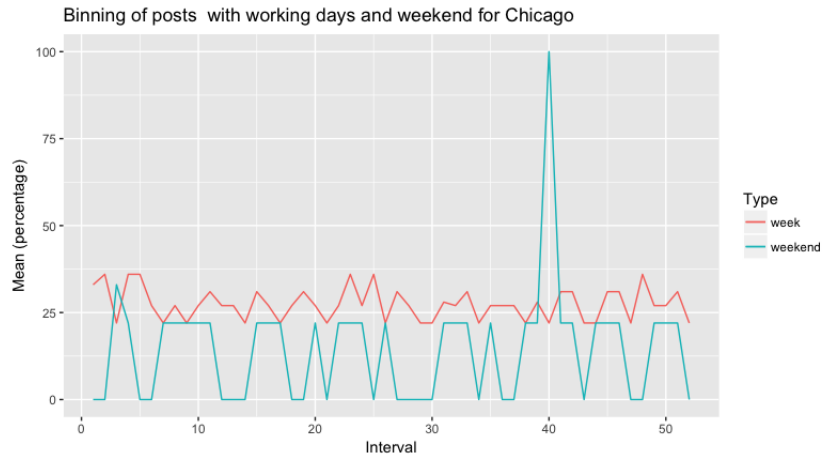


Figura 4.8: Agrupamento de *posts* para *Chicago*

instituição é notória a ausência de publicações até à vigésima semana, que diz respeito ao dia 20 de janeiro de 2016, que representa o término do período de leitura, prazo para a submissão dos trabalhos escritos dos alunos e início dos exames do semestre de outono. Na vigésima segunda semana, que corresponde ao período de 1 a 5 de fevereiro, a instituição alcança o pico correspondente à semana com uma taxa de 85%, diminuindo drasticamente na vigésima quinta, que diz respeito aos dias 22 ao 26 de fevereiro de 2016 (mais de 50%), volta novamente a aumentar significativamente o número de publicações, mantendo esse comportamento ao longo de todo o gráfico. O pico do gráfico é cumprido na trigésima oitava semana, nos dias 28 e 29 de maio que corresponde à cerimónia de bacharelato, e volta a diminuir, chegando mesmo à não existência de publicações.

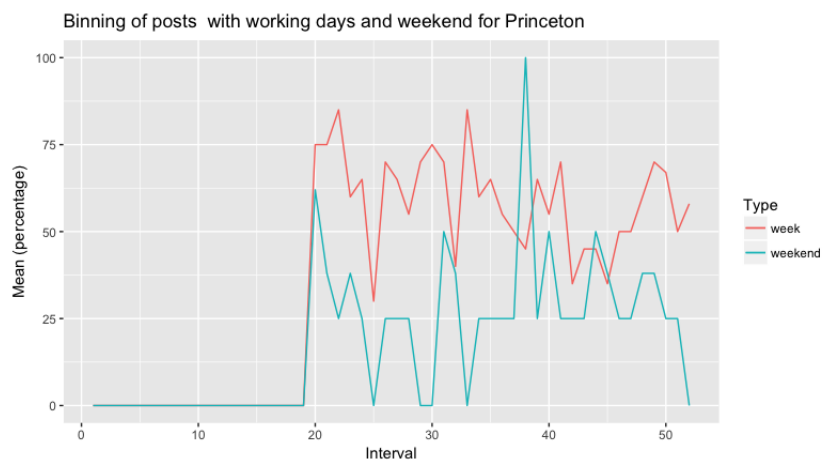


Figura 4.9: Agrupamento de *posts* para *Princeton*

Através da análise da figura 4.10, onde se verifica agrupamentos para sete dias, constata-se que Yale University é uma instituição ativa nas redes sociais, apresentando uma média diária de 2,431 *posts*. Ao longo da análise do gráfico é de notar a existência de um decréscimo do número de publicações no quarto ponto com uma percentagem de 68%, que diz respeito ao período entre os dias 19 e 25 de janeiro de 2016, que corresponde ao início do semestre de primavera, às

inscrições para todos os alunos e ao início das aulas. No ponto seis, com uma taxa de 100%, é atingido o pico do gráfico no intervalo do dia 2 ao dia 8 de fevereiro de 2016, de seguida, regista-se um decréscimo até ao oitavo ponto (com 82%) que representa o conjunto de dias do 16 ao 22 de fevereiro. Note-se a existência de uma quebra no número de publicações desde o ponto dezassete, constituído pelo agrupamento dos dias 19 ao 25 de abril, com uma taxa de 91%, até ao final do período envolvido no estudo. Nessa fase, na instituição, comemora-se o fim das aulas, a época de exames, a cerimónia de graduação e as férias escolares.

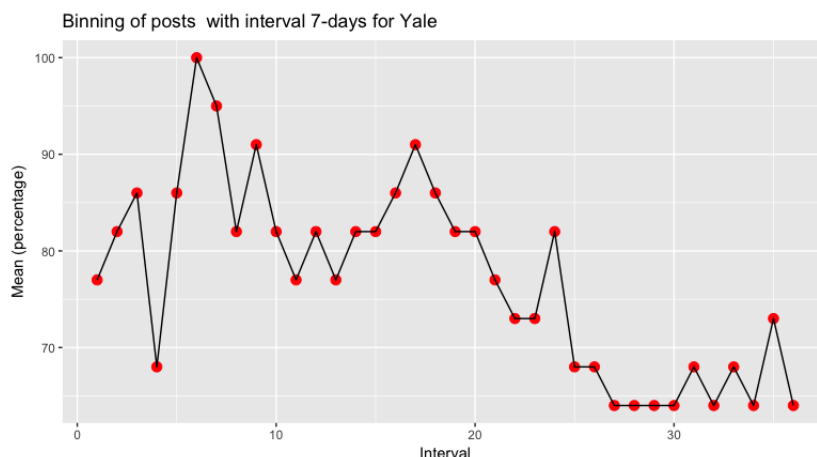


Figura 4.10: Agrupamento de *posts* para *Yale*

4.1.2 Interações ao longo do tempo

Na figura 4.11 foi adotada uma visualização com agrupamento de sete dias. Na análise de Harvard University verificamos que apesar de se verificar em grandes disparidades entre os valores médios a cada 7 dias de interações registadas na página de Harvard, é possível registar interações em todos os intervalos. É possível verificar ainda que a tendência geral de variação do número médio de interações 4.11 demonstra alguma correspondência com a tendência geral de variação do número de publicações 4.1. A correspondência não é exata no eixo do intervalo, porque os gráficos resultam de processos de *binning* distintos, embora os intervalos de 7 dias possam ser aproximados aos intervalos semanais. No entanto, é possível verificar que há uma tendência dos seguidores para aumentar o seu nível de participação quando a instituição também aumenta o número de publicações e vice-versa. É notório, no oitavo ponto, a existência de um pico, que apresenta uma percentagem de adesão de 100% aos *posts* realizados no conjunto de dias do 22 ao 28 de outubro de 2015, que diz respeito à publicação de *posts* direcionadas ao outono em Harvard. Outro ponto que também se destaca no gráfico é o ponto 20, com 81%, que corresponde ao conjunto de dias do 23 ao 29 de janeiro de 2016 que diz respeito ao início do trimestre de primavera, desde aí, nota-se uma ligeira descida no número de respostas obtidas até aos pontos vinte e sete, onde atinge os 11%. A partir daí, o número de respostas aumenta até ao ponto trinta e seis e trinta e sete, com 90% e 85%, respetivamente. Esses pontos são constituídos pelo intervalo de dias de 17 de maio até 29 de maio, que se referem ao fim dos exames, ao final do

ano letivo e ao dia de formatura.

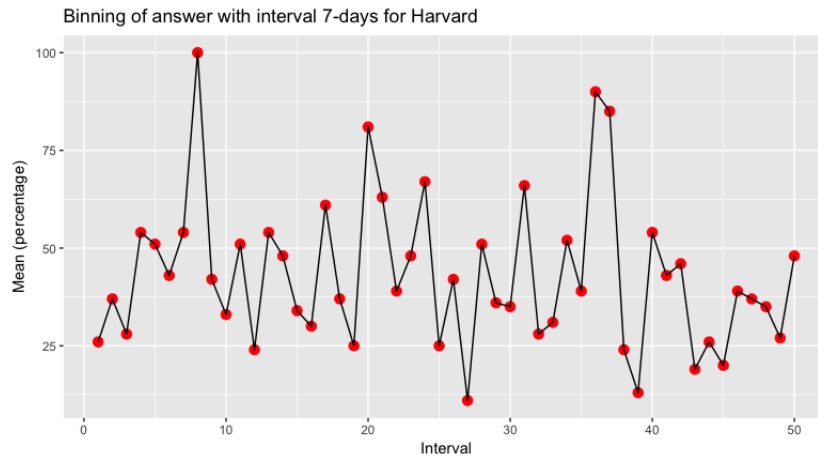


Figura 4.11: Agrupamento do número de respostas para *Harvard*

Através da análise da figura 4.12 e com a adoção de uma visualização de um *binning* de dias úteis e fins-de-semana, Stanford University obtém uma média de 837 respostas às publicações por dia. A adesão da comunidade da instituição, mostra preferência em *posts* realizados aos fins-de-semana, onde se regista, na décima quarta semana, uma taxa de adesão de 100% no dia 6 de dezembro de 2015, que diz respeito à véspera de exames de outono. A partir daí, há um decréscimo de publicações, cerca de 80%. No gráfico, é possível analisar mais dois picos aos fins-de-semana, sendo um na vigésima sétima semana com uma taxa de adesão de 73%, no dia 13 de março de 2015, onde se celebra o final do semestre de inverno e dia da morte de *Leland Stanford*¹ e início dos exames na faculdade de direito. A partir daí é notória uma diminuição, chegando às zero respostas obtidas. Na semana quarenta, no dia 4 de junho, é alcançado outro pico com 65%. O ponto que se destaca no número de respostas obtidas à semana é realizado entre os dias 13 e 17 de junho (quadragésima segunda semana) com 46% e o fim-de-semana, corresponde ao dia da formatura e alcança uma adesão de 30%

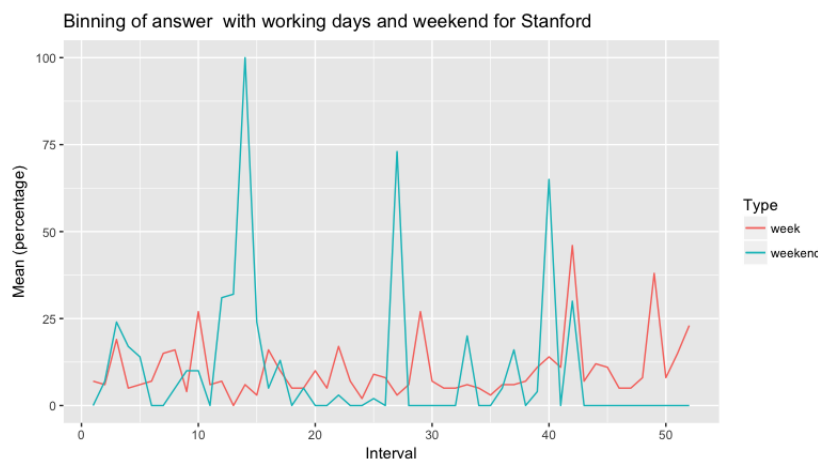


Figura 4.12: Agrupamento do número de respostas para *Stanford*

¹Stanford foi construída pelos pais de *Leland Stanford*, em sua honra

A figura 4.13 apresenta a atividade dos seguidores de MIT através do uso de um *binning* de dias úteis e fins-de-semana, apresentando uma média de 1380 respostas diárias obtidas. Nesta instituição destaca-se o número de interações obtidas à semana, na vigésima terceira semana, com uma taxa de 100%, que diz respeito aos primeiros dias do trimestre de primavera (intervalo de dias 8 a 12 de fevereiro de 2012) e verifica-se uma discrepância de 90% para a semana seguinte. É notório outro pico no gráfico, referente às semanas trinta e nove e quarenta (28 e 29 de maio e 4 e 5 de junho) com 60% e 57% onde se celebra a data limite da pré-inscrição de verão, o dia de formatura, cerimónia de doutoramento e o primeiro dia da sessão regular de verão. O número de respostas obtidas ao fim-de-semana varia entre os 3% e os 44%.

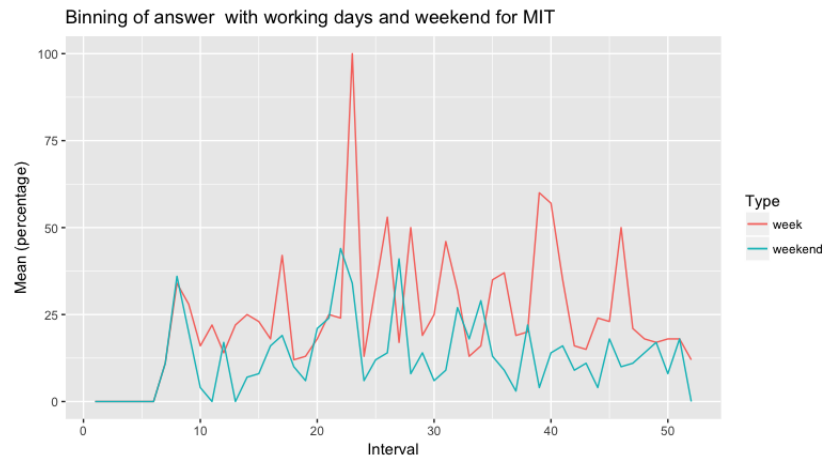


Figura 4.13: Agrupamento do número de respostas para *MIT*

A figura 4.14 apresenta a atividade desenvolvida pelos seguidores de University of Cambridge através de um *binning* de dias úteis e fins-de-semana. É importante frisar que a sua comunidade apresenta uma atividade de 2377 respostas por dia. É possível verificar que a instituição apresenta ausência de resposta na maior parte dos fins-de-semana, sendo os outros com pouca adesão. Durante a semana, a adesão por parte do público-alvo é maior, destacando-se dois picos, o primeiro na segunda semana com 100%, que diz respeito às férias escolares antes do ano letivo

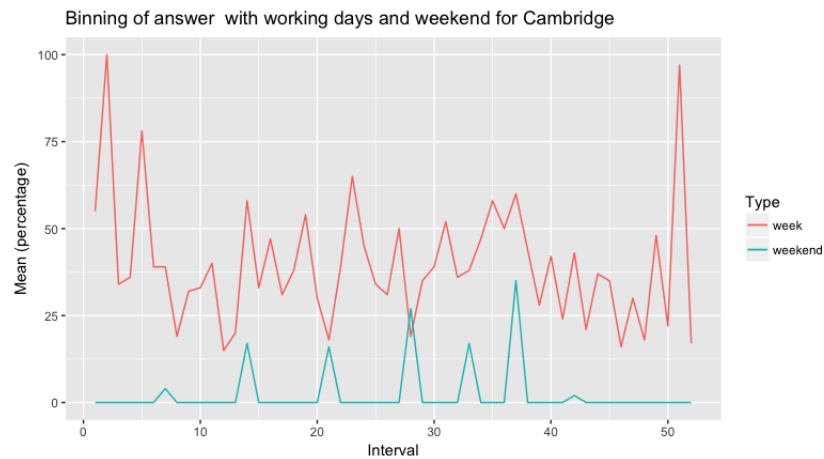


Figura 4.14: Agrupamento do número de respostas para *Cambridge*

2015-2016 e o outro pico com 97% na semana cinquenta e sete nas vésperas de iniciar o ano letivo 2016-2017. Note-se que são raras as vezes que o número de respostas realizadas em *posts* publicados à semana são inferiores a 20% e que comparado com o número de publicações 4.4, quando a quantidade de *posts* é maior o número de respostas é menor.

No gráfico de *binning* de sete dias da figura 4.15 estão refletidas as publicações de University of Oxford que apresenta uma média de 2880 respostas diárias. O primeiro momento de destaque existente no gráfico refere-se ao ponto dezasseis com uma taxa de respostas obtidas de 99% e refere-se ao período de férias de natal composto pelos dias 21 a 27 de dezembro de 2015. Com uma taxa de adesão de 100%, o ponto vinte e um, composto pelos dias 10 a 16 de março de 2016, onde se efetuam as reuniões de negócios e comissões. É possível verificar que do ponto vinte e sete para o vinte e oito existe uma queda em relação ao número de respostas obtidas pela instituição. É notório no ponto trinta e dois a ausência de gráfico, isso deve-se ao fato de no dia 23 de abril o *Facebook* não registar qualquer informação relativamente às *reactions*.

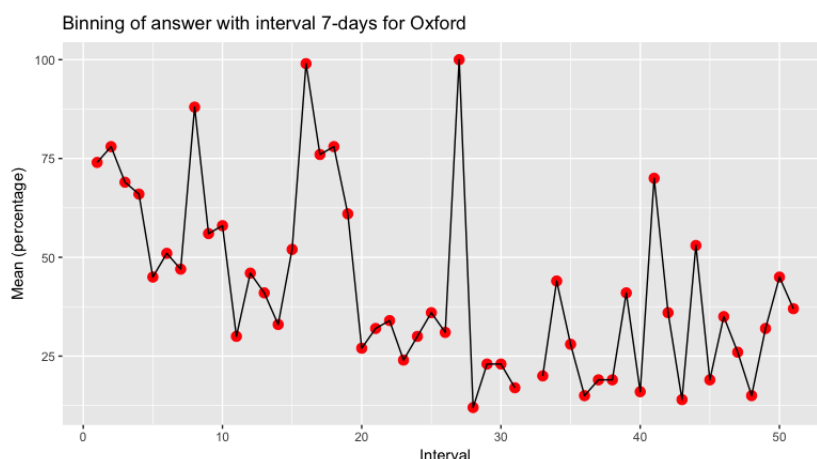


Figura 4.15: Agrupamento do número de respostas para *Oxford*

Na figura 4.16 é observável o registo das publicações de Columbia University ao longo de um ano letivo. Regista-se uma média diária de 563.5 respostas obtidas às publicações realizadas por Columbia. Esta instituição é a única que regista ausência de respostas tendo efetuado publicações nesse período. Através do gráfico, é possível concluir que a comunidade de Columbia é pouco participativa, visto que o número de respostas recolhidas, tanto em dias úteis como fins-de-semana não ultrapassam os 25% com a exceção do trigésimo nono fim-de-semana, onde regista uma taxa de adesão de 100%, período antes do registo de verão para o trimestre de outubro e o dia de honrar *Muhammad Ali* ² (4 de junho de 2016). Durante a semana, é também registado um pico com 51% de adesão por parte da sua comunidade, que diz respeito à trigésima sétima semana, composta pelos dias 16 ao 20 de maio de 2016, que regista o regresso dos estudantes do ensino profissional.

A figura 4.17 apresenta-nos uma análise das respostas obtidas aos *posts* realizados por UC Berkeley utilizando um *binning* de dias úteis e fins-de-semana. Analisando o gráfico, é possível

²lutador de boxe americano

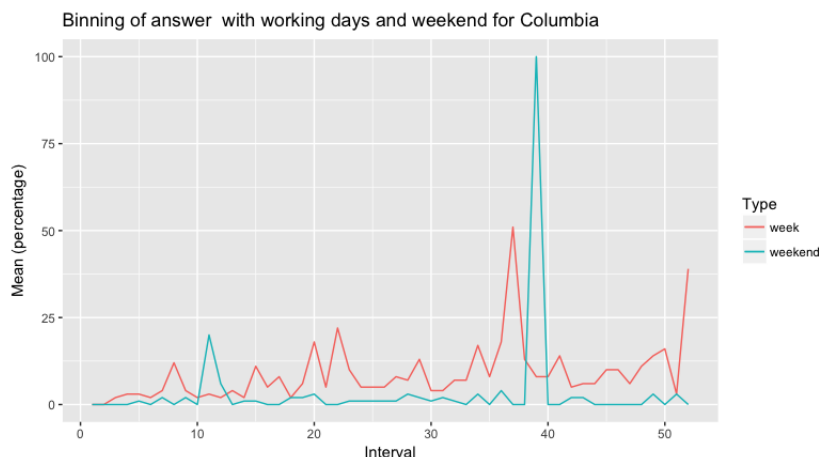


Figura 4.16: Agrupamento do número de respostas para Columbia

verificar que são obtidas mais respostas da comunidade, à semana. É durante o fim-de-semana que são registados os picos de adesão, sendo eles na quadragésima quinta semana, dia 17 de julho, com 100%, onde os alunos da instituição ajudaram os sem-abrigo através da *Suitcase Clinic* e o dia da formatura com 67% no dia 14 de maio de 2016, na trigésima sexta semana. Com a análise do gráfico é notório que entre a semana quarenta e cinco e quarenta e seis existe uma discrepância de, aproximadamente, 80%. Com ajuda do gráfico 4.7, é possível verificar que a instituição recebe menos respostas do que publica.

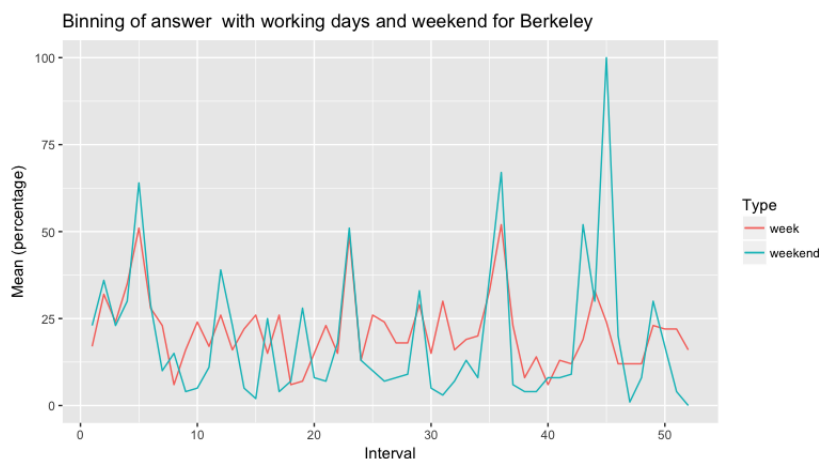


Figura 4.17: Agrupamento do número de respostas para Berkeley

Através da análise da figura 4.18 constatamos, através de um *binning* de dias úteis e fins-de-semana, que University of Chicago recolhe mais respostas do que o número de publicações que realiza. É notório que a instituição reúne um maior número de adesão com *posts* realizados à semana, mas é com o ponto onze, que o fim-de-semana se sobressai com uma adesão de 100%, com uma mensagem sobre os atentados ocorridos em Paris no dia anterior. Do ponto onze ao ponto doze é sentida uma discrepância de 100% devido à ausência de *posts*. O segundo valor mais alto obtido pela comunidade regista-se no período de leitura universitária, no dia 31 de maio, 1, 2 e 3 de junho, com 65%,

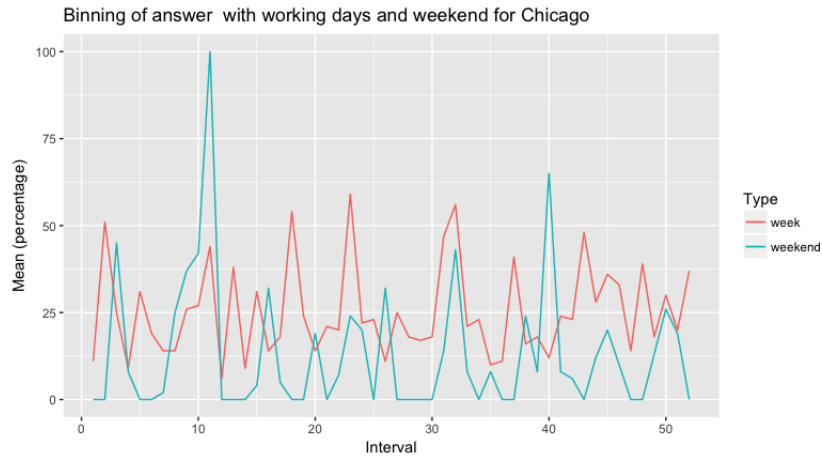


Figura 4.18: Agrupamento do número de respostas para *Chicago*

A figura 4.19 apresenta a atividade da página de Princeton University através do uso de um *binning* de dias úteis e fins-de-semana. Neste gráfico notam-se grandes oscilações sobretudo aos fins-de-semanas e que a média de respostas obtida é de 503.9 e à semana é de 806.4 respostas. É possível verificar que no ponto vigésimo sétimo onde se inicia o trimestre de primavera (12 de março), há 100% de adesão por parte da comunidade da instituição. Durante a semana é também registado um pico, com 98% de adesão, nos dias 14 ao 18 de março (ponto vinte e oito) onde os alunos assistem à primeira semana do trimestre de primavera. O número de publicações realizadas nos dias úteis é maior que o número de resposta por parte da comunidade, o mesmo não se passa ao fim-de-semana.

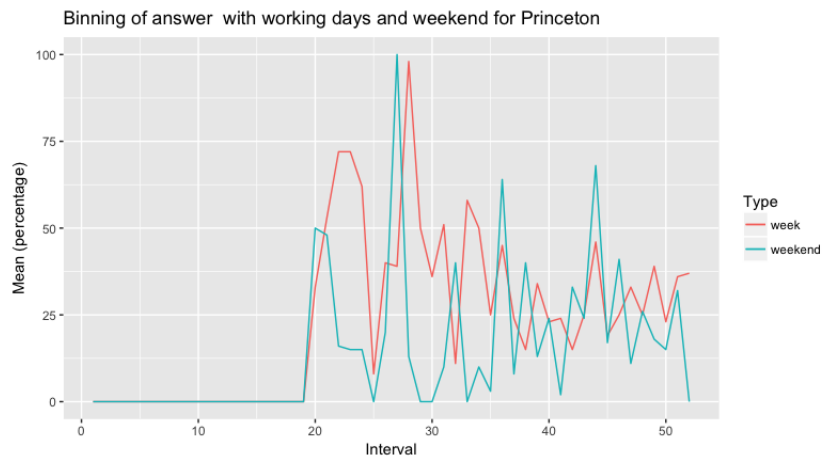


Figura 4.19: Agrupamento do número de respostas para *Princeton*

Na figura 4.20 é visível um *binning* de dias úteis e fins-de-semana referente às respostas obtidas de Yale University. Nesta instituição é notória a ausência de atividade até à décima oitava semana, 29 de dezembro de 2015, registando a partir desse momento uma comunidade bastante participativa com médias de 801,1 à semana e 754,1 ao fim-de-semana. A partir do gráfico é possível registar a existência de quatro pontos cruciais, sendo três deles ao fim-de-semana, no ponto trinta e um (2 e 3 de abril) com 85%, no ponto trinta e cinco (30 de abril e 1 de maio)

com 100%, em que os alunos se encontram no período de leitura e o ponto cinquenta e um (20 e 21 de agosto) com 74%. Durante a semana, também se nota a existência de um pico no ponto cinquenta e um, composto pelos dias 15 ao 19 de agosto, com 85%.

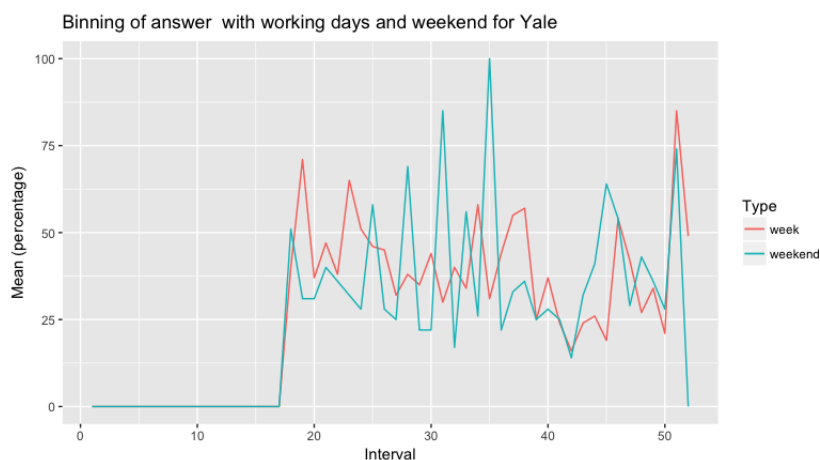


Figura 4.20: Agrupamento do número de respostas para *Yale*

4.2 Análise de desempenho global por trimestres

Analisando a informação recolhida, sente-se a utilidade de estudar o desempenho praticado pelas instituições a nível trimestral. Visto isto, os dados são separados por universidade e por períodos de tempo de forma trimestral. Para esta análise apenas são consideradas variáveis do tipo numérico e os nomes das instituições, para assim perceber a que universidade os resultados dizem respeito. Para esse efeito, todos os valores de *NA*'s são convertidos para zeros, para assim facilitarem a obtenção de dados e agrupados de forma a ser perceptível o número de *posts* por instituições e a quantidade de respostas obtidas pelas publicações.

Nos seguintes gráficos podemos analisar o desempenho global por trimestres, uma vez que mostram a performance das universidades a nível de respostas e a nível de publicações. Sendo que cada cor corresponde a uma instituição, é esperado que sejam comparadas as diferentes universidades quanto à relação existente entre o número de respostas e o número de *posts*. Quanto maior é o número de respostas maior é a interação que a comunidade mantém com a instituição, por outro lado, quanto menor é esse valor mais intiva é o público-alvo. O mesmo se passa em relação ao número de *posts*.

No primeiro semestre do ano, composto pelos meses de setembro a novembro, apresentam unicamente oito instituições, visto que Princeton University e Yale University não tem qualquer atividade neste período. Como se observa na figura 4.21, University of California, Berkeley é a instituição com mais publicações realizadas, por outro lado, Harvard University manifesta uma grande quantidade de interação comparadas às outras instituições. Stanford University, é realçada pela falta de publicações e de respostas por parte dos seguidores. No entanto, Columbia University demonstra uma comunidade pouco ativa, visto que possui uma alta taxa de publicação

comparada com a taxa de retorno.

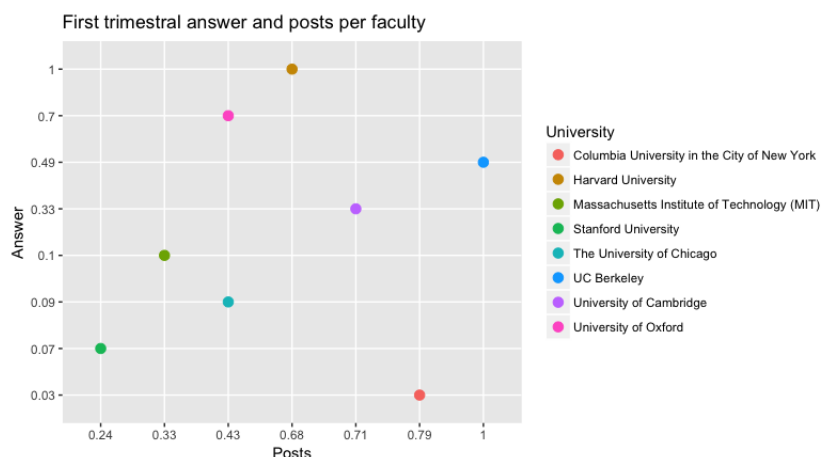


Figura 4.21: Desempenho global no primeiro trimestre do ano letivo

No gráfico referente ao segundo trimestre do ano, representado na figura 4.22, composto pelos meses de dezembro de 2015 a fevereiro de 2016, é notório que Harvard University mantém a mesma posição que no trimestre anterior. Neste trimestre, as dez instituições já fazem parte da composição do gráfico. Columbia University é a instituição com mais publicações, mas, em contrapartida, obtém menos interações por parte dos seus seguidores. Stanford University, continua a ser a instituição com menos *posts* divulgados. Por outro lado, University of Oxford diminui os valores alcançados, mas permanece na segunda posição a nível de respostas alcançadas.

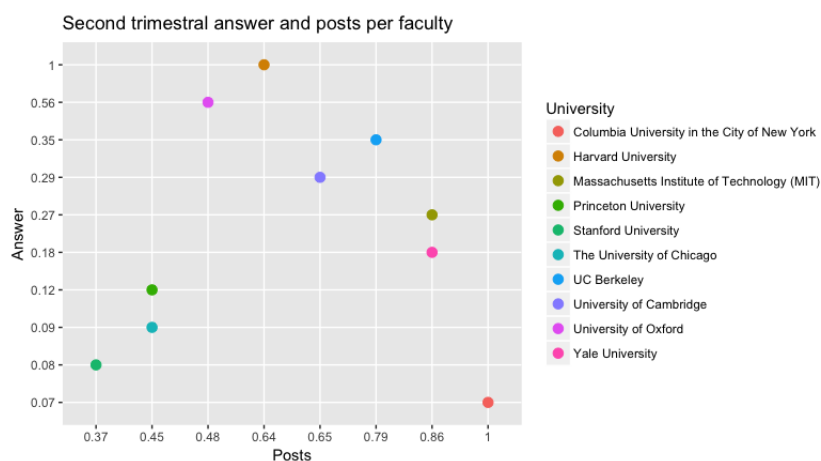


Figura 4.22: Desempenho global no segundo trimestre do ano letivo

Os meses de março a maio, compõem o terceiro trimestre do ano. Neste trimestre, Harvard University preserva a mesma posição, diminuindo um pouco a quantidade de material publicado. University of Oxford diminui a quantidade de publicações, contudo o número de resposta dos seus seguidores também diminui. Do mesmo modo que Stanford University conserva a mesma localização, sofrendo uma diminuição dos dois fatores. Yale University manifesta um grande aumento a nível de publicações, ocupando assim a posição com maior número de *posts*.

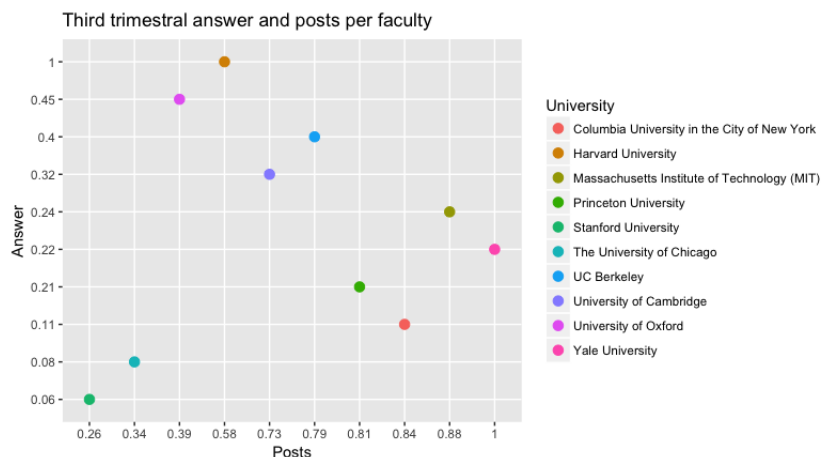


Figura 4.23: Desempenho global no terceiro trimestre do ano letivo

Finalmente, o último trimestre do ano é composto pelos meses de Junho a Agosto, este comprova que Harvard University continua a ser a instituição que mais respostas conquista. Stanford University, Yale University e University of Oxford preservam a mesma posição, com alterações mínimas nos valores apresentados. Columbia University continua a receber uma baixa taxa de retorno, dado a quantidade de publicações.

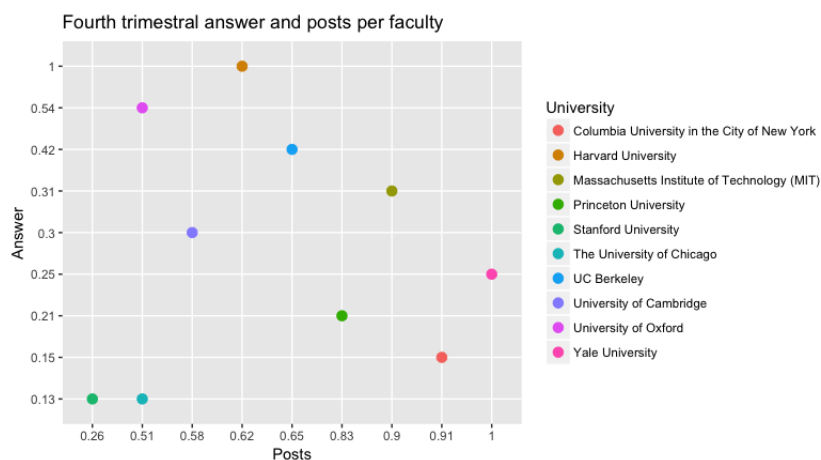


Figura 4.24: Desempenho global no quarto trimestre do ano letivo

Com isto é aceitável concluir, que de uma forma geral, as instituições divulgam mais informação nas redes sociais em finais de trimestres, épocas de exames e no dia da cerimónia de formatura. A comunidade das instituições é bastante ativa em publicações realizadas nos finais e inícios dos trimestres, cerimónia de formatura e em dias marcantes para a instituição, como é o caso de Stanford University e University of Oxford em que ambas têm um dia dedicado á sua fundadora e fundação da instituição, respetivamente.

Relativamente ao desempenho por trimestres das instituições, Harvard University obtém maior resposta por parte dos seguidores em todos os trimestres analisados, porém a quantidade de *posts* divulgados diminui ao longo do ano. Na University of Oxford, ao longo do ano constata-se

um aumento, ligeiro, de número de publicações e uma diminuição de retorno da sua comunidade. University of Cambridge demonstra um comportamento irregular com uma preservação do número de respostas. Em contrapartida University of California, Berkeley vai diminuindo a utilização das redes sociais ao longo dos trimestres considerados. Columbia University, Princeton University, Yale University e Massachusetts Institute of Technology, ao longo dos trimestres, aumentam o número de publicações divulgadas e a sua comunidade aumenta a capacidade de resposta.

Capítulo 5

Análise de conteúdo

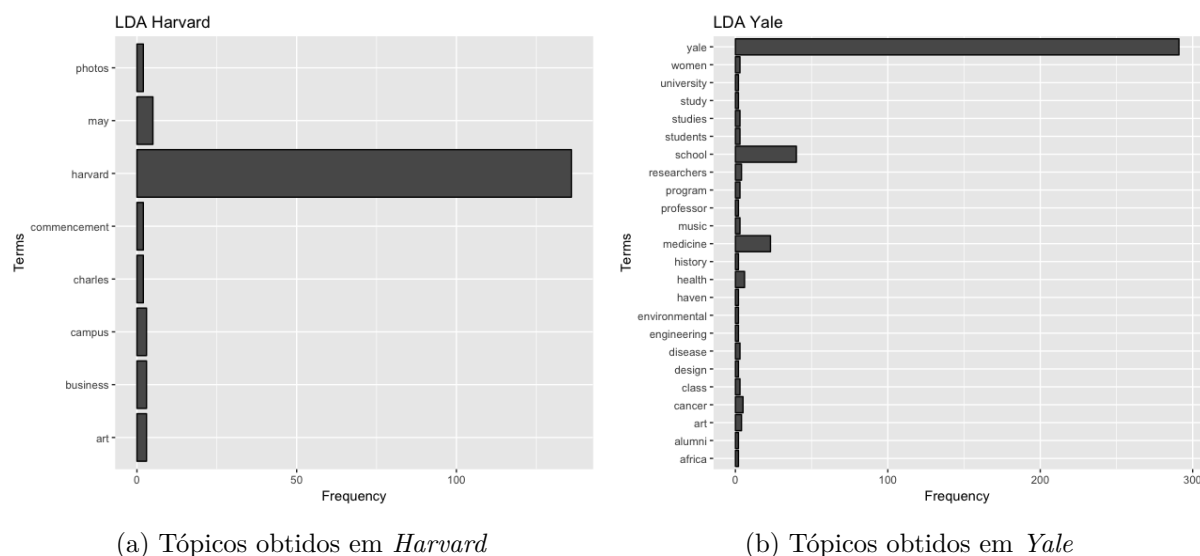
Neste capítulo é narrado todo o processo orientado ao estudo do conteúdo existente nas publicações executadas pelas dez melhores instituições de ensino superior do mundo. Inicialmente, é desenvolvido um procedimento para analisar a importância do conjunto de palavras para o *post* realizado. Por último, é retratado um mecanismo de detecção de tópicos existentes nas publicações em causa.

5.1 Análise da importância das palavras para o *post*

Tendo em conta a relevância dos *posts* no presente estudo, é executado um processo que identifica a importância das palavras que compõem a mensagem. Para tal, é utilizada uma medida estatística utilizada em DM, *Term Frequency - Inverse Document Frequency* mais conhecido por TF-IDF.

Com o propósito de realizar o processo para as diferentes universidades inseridas no estudo, é realizado um conjunto de regras que nos permite a concretização da tarefa. No começo, o *post* é tratado como foi referido no capítulo 3, à exceção das expressões retiradas nas *stopwords*, visto que cada instituição usa palavras com pouca relevância para o estudo que não estão contidas nessa lista, e da realização de redução de *stemming* visto que com isto, não é possível perceber quais as palavras com mais importância, mas sim quais as raízes das palavras. Imediatamente, é construída uma matriz termos-documentos com o peso do TF-IDF, ou seja, é calculado a frequência das palavras e o peso das mesmas são diminuídos, imediatamente, como mencionado no capítulo 2.

Para uma visualização dos termos, é elaborado uma *wordcloud* que contribui para uma melhor visualização daqueles que têm mais importância, onde o tamanho da palavra, na imagem, indica o seu interesse. Assim sendo, para todas as universidades são utilizadas os mesmos parâmetros cuja regularidade mínima das palavras seja de um e cada gráfico é composto, no máximo, por cem palavras.

Figura 5.11: Agrupamento das instituições *Harvard* e *Yale*

5.2 Análise dos tópicos das mensagens

Posteriormente à identificação das palavras com mais importância, segue-se a identificação dos tópicos existentes nas mensagens que contenham as palavras mais importantes. Simplificando, o processo de deteção dos tópicos das mensagens foi realizado tendo em conta as palavras de grande importância. Para este processo, apenas expressões identificadas em cima e que possuam uma percentagem de importância a cima dos 30% a 50%, são tidas em conta. Inicialmente, são recolhidos todos os *posts* que contenham as palavras identificadas pelo processo TF-IDF, essas mensagens são todas elas separadas por instituições guardadas num *dataset* distinto. Posto isto, o método de deteção de tópicos, LDA, é aplicado sobre esse *dataset* que contem as mensagens tratadas e utilizadas na secção acima referida.

Como se verifica na ilustração 5.11a, Harvard University capta a atenção dos seus seguidores com termos relacionados à imagem e seus eventos, usando termos como *harvard*, *commencement*, *may*, *photos*, mas não exclui termos como *campus*, *bussiness*, *art* e *charles*. Note-se que há uma grande diferença entre *harvard* e os restantes tópicos. Em Yale University, além de se destacar expressões como *yale*, a instituição utiliza termos no âmbito do ensino, como é o caso de *medicine*, *students*, *studies*, *study* e *engineering*, como se verifica 5.11b. Estas duas instituições têm em comum o ênfase dados às áreas de educação, à imagem e às áreas de estudo, com expressões como *music*, *medicine*, *history*, *art* e *business*, demonstrado no ilustração¹ 5.11.

Através da ilustração 5.12b verifica-se que Massachusetts Institute of Technology apresenta uma quantia de tópicos associados ao ensino, *physics* e *students*, não abandonando o tópico que permite enaltecer a imagem da instituição em causa, MIT. Um tópico que se evidencia por nome diferente é *hyperloop*, que diz respeito a competições relacionadas com a tecnologia,

¹as imagens estão disponíveis em maior dimensão em apêndice

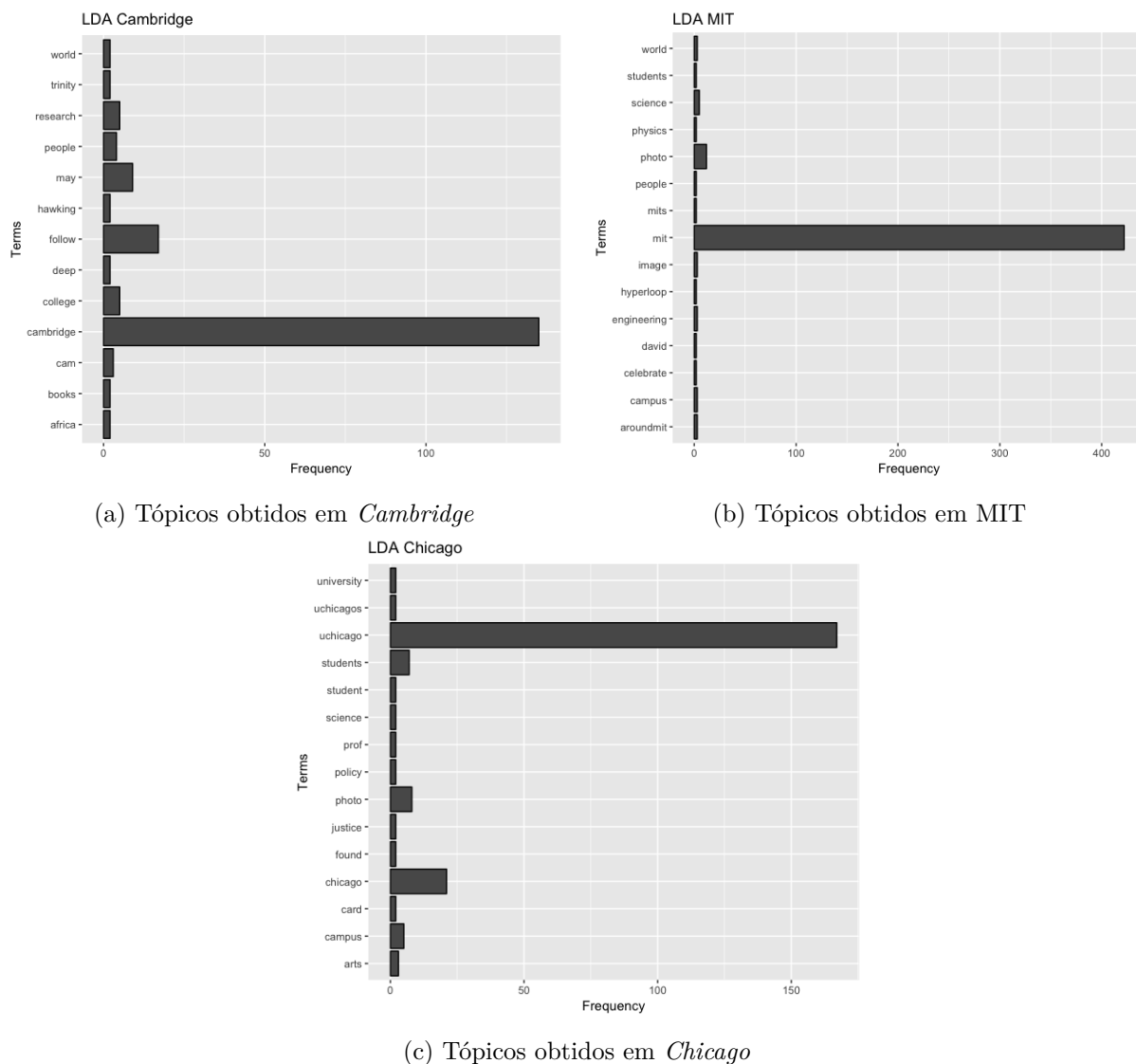
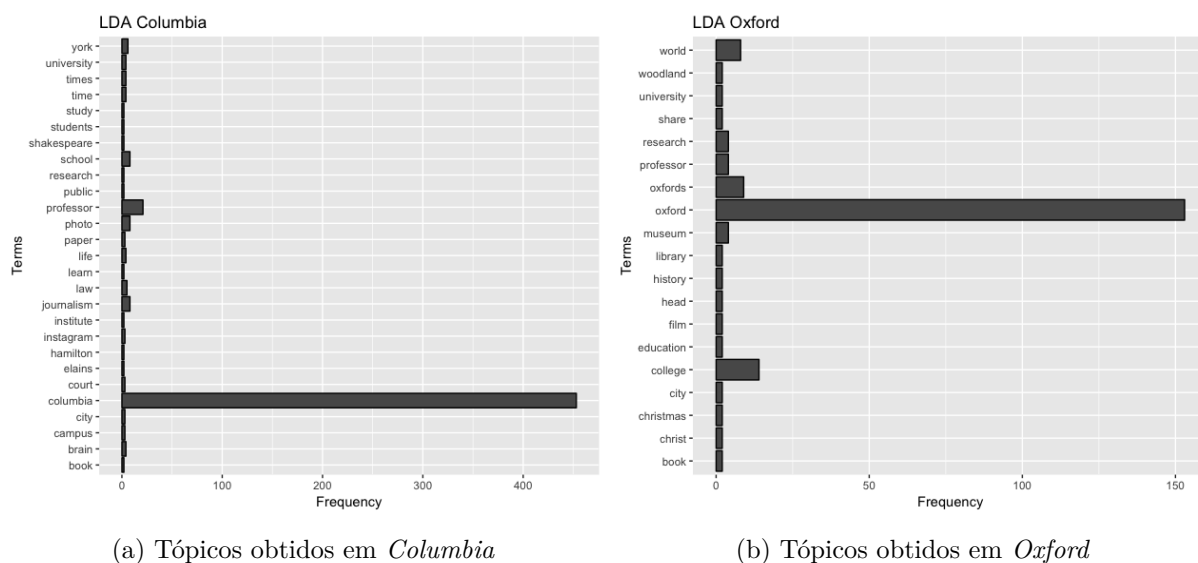


Figura 5.12: Agrupamento das instituições *Cambridge*, MIT e *Chicago*

onde, enquanto participavam, os seus estudantes obtinham condecorações. Em University of Cambridge destaca-se o termo *cambridge* em que refere à instituição em causa. Através da ilustração 5.12a, é permitido identificar termos diferentes, como *trinity* que identifica uma parte constituinte da universidade. Outro termo que se destaca nesta instituição é *hawking*, refere-se ao conhecido professor cientista *Stephen Hawking*, que neste momento é diretor de pesquisa do Departamento de Matemática Aplicada e Física Teórica e fundador do Centro de Cosmologia Teórica da Universidade de *Cambridge*. Em University of Chicago dos tópicos representados na ilustração 5.12c, aquele que mais destaca é referente à instituição em causa. Nesta instituição, tópicos relacionados com a educação possuem um grande valor, como é o caso de *students*, *science* e *arts*. Outros dois tópicos que a universidade destaca são *policy* e *justice*. A MIT, *Cambridge* e *Chicago* têm em comum o fato de evidenciar áreas como a instituição, MIT, *cambridge* e *chicago*, a educação com expressões como *college*, *students* e *prof* e as redes sociais usando os tópicos *follow*, *cam* e *photo*. Em MIT e *Chicago* dão ênfase à área da globalização através da expressão

Figura 5.13: Agrupamento das instituições *Columbia* e *Oxford*

world, como se verifica na ilustração² 5.12

Com a ilustração apresentada em 5.13b, é visível a dimensão do tópico *oxford*, destacando-se também a época natalícia com a utilização do termo *christmas*. Uma expressão pouco comum no nosso vocabulário, mas que surge na imagem, é *woodland* que é o nome dado a um departamento da instituição. Os outros tópicos apresentados andam sempre à volta do mesmo tema, como é o caso de *university*, *research*, *college*. Na Columbia University, os tópicos são relacionados com a instituição, sendo o nome da instituição, *columbia*, que apresenta grande destaque. O nome de um departamento da instituição destaca-se visto ser o mesmo de um poeta mundialmente conhecido, *shakespeare*. Pela visualização da ilustração 5.13a denota-se a existência de outro termo não muito popular é *hamilton*, nome dado a um edifício existente dentro do *campus*. Estas duas instituições têm em comum o fato de usufruir de áreas como a imagem da instituição, a educação, as redes sociais com expressões como *share*, *photo* e *instagram* e pelo ênfase aos departamentos/áreas da instituição como *shakespeare*, *hamilton*, *law*, *library*, *museum*, *jornalism*, como se verifica na ilustração³ 5.13.

Stanford University dá ênfase a tópicos nas suas publicação relacionados com o desporto, com expressões como *football* e *gostanford*. Esta instituição, como ilustra em 5.14c, dá mais importância aos tópicos relacionados com a imagem e com os estudantes *stanford*, através dos termos com o maior número de frequência e *students*, que é o segundo tópico mais usual. Como se demonstra na imagem 5.14b, University of California, Berkeley expõe uma quantia considerável de termos, salientando *Kristine* que corresponde ao nome de uma estudante que se destaca pelos diversos prémios recebidos. Nesta instituição, salienta-se o tópico *clinton* que diz respeito ao ex-presidente *Bill Clinton*. O tópico utilizado em *UC Berkley* que recebe mais ênfase é *berkeley* seguindo-se de *gobears*. A Princeton University, apresenta uma seleção em tópicos destinados à

²as imagens estão disponíveis em maior dimensão em apêndice

³as imagens estão disponíveis em maior dimensão em apêndice

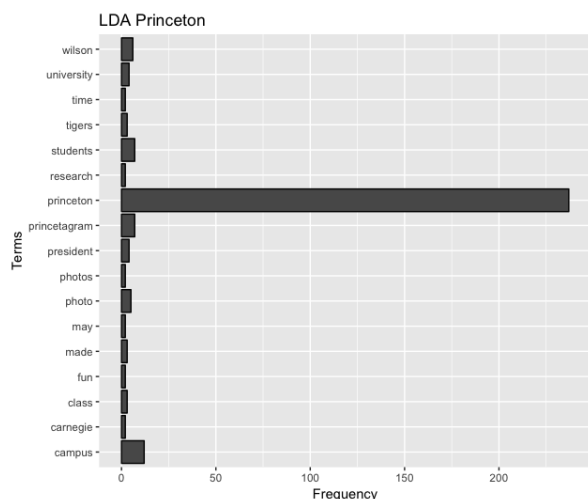
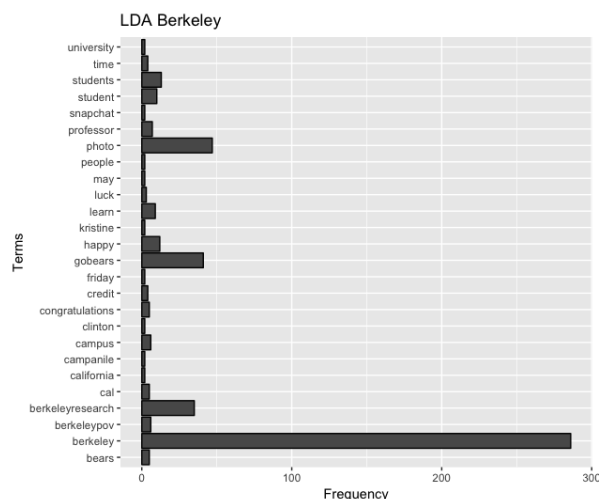
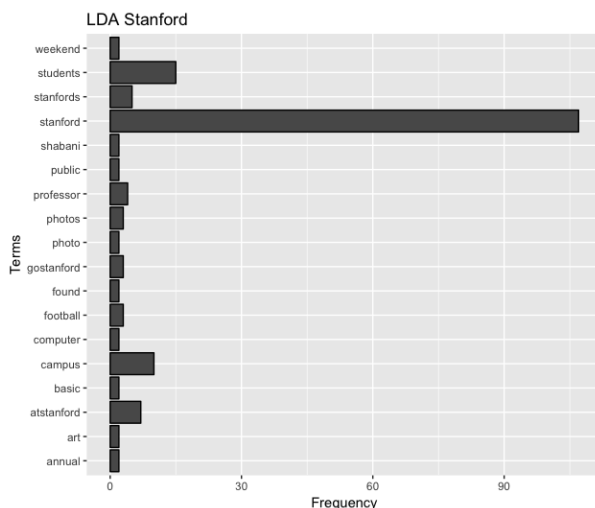
(a) Tópicos obtidos em *Princeton*(b) Tópicos obtidos em *Berkeley*(c) Tópicos obtidos em *Stanford*Figura 5.14: Agrupamento das instituições *Princeton*, *Berkeley* e *Stanford*

imagem da instituição, como é o caso de *princeton* que é a expressão mais frequente. Na imagem 5.14a, é ilustrada a dedicação ao ensino, utilizando expressões como *students*, *class*, e ao futebol, *tigers*. Um tópico que facilmente se diferencia é *carnegie*, que diz respeito a uma parte do *campus* criada para acolher a vida selvagem.

No agrupamento constituído pela ilustração⁴ 5.14, estão presentes que instituições possuem nas suas publicações de tópicos muito similares. Essas expressões salientam áreas como desporto, *tigers*, *bears*, *gobears* e *football*, a imagem da instituição com expressões como *princeton*, *stanford* e *berkeley*. Estas instituições, também dão ênfase à educação e às redes sociais com o uso de palavras como *professor*, *students* e *photo*. A UC Berkeley e Princeton University valorizam a área da felicidade com as expressões *happy* e *fun* e a área da pesquisa com o tópico *research*.

⁴as imagens estão disponíveis em maior dimensão em apêndice

5.3 Conclusões

5.3.1 Importância das palavras

Por estas razões, é possível concluir que todas as instituições dão grande ênfase ao nome e à sua imagem, à exceção de Yale University que não dedica tantos recursos para enaltecer o seu nome. É também notório que todas as instituições envolvidas no estudo fazem referência à educação, tendo UC Berkeley e Stanford University não dado tanta importância a esses temas, ao contrário de Yale University, em que as palavras com mais importância são direcionadas para esta área. No domínio do desporto, destacam-se a UC Berkeley e Stanford University, visto que a primeira palavra mais importante é direcionada à equipa de futebol. A Princeton University também publica de forma a dar ênfase à prática desportiva mas com menos ênfase. É de notar que Massachusetts Institute of Technology, Harvard University, UC Berkeley, University of Cambridge, University of Columbia, University of Chicago e Princeton University referem as fotos e redes sociais nas suas publicações, sendo Cambridge a instituição que atribui mais ênfase (primeira palavra), seguido de MIT e Princeton onde é a segunda palavra. Há ainda instituições que dão muita importância à felicidade, como é o caso de Massachusetts Institute of Technology e University of Oxford onde é a terceira palavra mais importante. Princeton University, UC Berkeley, University of Chicago também destacam a felicidade. Todas as instituições, à exceção de Harvard University, Stanford University e UC Berkeley, destacam a pesquisa mas sem grande intensidade. Por fim, University of Cambridge, University of Oxford são as únicas instituições que fazem referência ao mundo.

5.3.2 Tópicos das mensagens

Em suma, é possível salientar que todas as instituições envolvidas no estudo dão ênfase a tópicos relacionados com a imagem da universidade e à educação. A Columbia University, University of Oxford, Princeton University, UC Berkeley, Stanford University, MIT, University of Chicago e University of Cambridge utilizam publicações que dão ainda destaque às redes sociais por elas utilizadas. Stanford University, UC Berkeley e Princeton University são as únicas instituições que aplicam nas suas publicações tópicos relacionados com a prática desportiva entre as comunidades académicas. É notório que Princeton University e UC Berkeley valorizam tópicos relacionados com a felicidade e University of Cambridge e MIT dão uma certa importância à globalização.

Capítulo 6

Conclusões

6.1 Resumo do Trabalho

Tendo em conta a interação entre as instituições de ensino superior e a comunidade, é possível afirmar que o *Facebook* é a rede social mais significativa devido ao elevado número de utilizadores, à sua distribuição pela rede mundial e aos vários modos e formas de interagir com o público.

De forma a perceber como é que as instituições de ensino superior melhor classificadas no *ranking* mundial interagem com o público, foi desenvolvida esta dissertação, analisando a informação disponibilizada pelo *Facebook*. Neste âmbito, foram considerados o número de *posts*, o número de interações e as mensagens.

Numa primeira etapa identificaram-se as dez melhores instituições pertencentes ao *ranking* mundial, CWUR, e verificou-se qual a rede social mais utilizada por elas. Seguidamente, recolheu-se toda a informação relativa a um ano letivo completo da cada instituição e procedeu-se ao tratamento da mesma. Posteriormente, realizou-se uma análise dos dados, utilizando métodos de simplificação e compreensão da informação recolhida.

Por fim, procedeu-se ao estudo do comportamento e desempenho das instituições e analisar-se as palavras e tópicos mais importantes utilizados e empregues nas mensagens de cada instituição.

6.2 Conclusões Gerais

- É possível concluir que Columbia University é a instituição que ao longo do ano realiza mais publicações e a que emprega um maior número de palavras. Por outro lado, Stanford University recolhe um menor número de respostas ao longo do ano. É a instituição que menos publica e que utiliza menos palavras nos *posts* realizados. Também Columbia University apresenta um reduzido número de respostas (9%).
- Harvard University é instituição que capta mais interesse por parte da comunidade, visto

que o número de respostas é mais elevada. Harvard University apresenta 100% de respostas ao longo do ano, enquanto que University of Oxford apresenta apenas 60%, tendo as outras instituições um menor número de respostas.

- Constatamos nas visualizações de dias úteis e fins-de-semana que as instituições realizam mais publicações durante a semana. Contudo não são essas publicações que obtêm mais interações, uma vez que se verificam alguns fins-de-semana com mais interações do que alguns dias úteis. Em University of Chicago e University of Cambridge verificamos que quando o número de *posts* aumenta o número de respostas realizadas pela comunidade é menor e vice-versa.
- Em relação ao desempenho global por trimestres, é notório que Harvard University é a instituição com uma maior capacidade para interagir com a comunidade. Esta instituição não é a que apresenta o maior número de publicações em nenhum dos trimestres, visto que Yale University, Columbia University e UC Berkeley lideram sempre em relação ao maior número de *posts*. Contudo, Harvard University é a instituição que regista sempre um maior número de respostas por parte da sua comunidade.
- Verifica-se que todas as universidades dão relevo ao nome, variando entre 100% e os 63%, ao passo que Yale University apresenta 0%. Por outro lado, Yale University dá relevo ao ensino, apresentando 100%, tal como Columbia University (100%). Stanford University, UC Berkeley e Princeton University são as únicas instituições que dão ênfase ao desporto apresentando 100%, 100% e 49%, respetivamente.
- Tendo em conta os resultados obtidos, constata-se que o *Facebook* é vital para a interação entre a maioria das instituições e as suas comunidades. Contudo é visível que para outras como Stanford, esta rede social não assume grande importância.
- Foi possível verificar que os tópicos mensagens variam em função da instituição e que não há relação linear entre as publicações aos fins-de-semana e dias úteis nas universidades envolvidas neste estudo.

6.3 Trabalho Futuro

Tendo em conta a constante utilização do *Facebook* como meio de comunicação entre as instituições e o público-alvo, verifica-se uma necessidade cada vez maior de construir presença na rede social em questão. Desta forma, pretende-se detetar os tópicos por trimestres ou períodos temporais para perceber se há assuntos que dominam mais numa determinada altura do que em outra.

Tenciona-se proceder à realização do estudo de comportamento na parte de respostas em que em vez de se realizar uma soma básica, se faria uma soma ponderada entre os *comments*, *likes* e *reactions*. Ambiciona-se também utilizar o NLP, que ajuda a perceber qual o significado das palavras, tendo em conta o contexto em que está inserido.

Bibliografia

- [1] Alexandra Isabel Magalhães Alves. Modelo de representação de texto mais adequado à classificação, 2010.
- [2] William B. and John M. Trenkle. N-gram-based text categorization.
- [3] Kalpana Chauhan and Anandan Pillai. [Role of content strategy in social media brand communities: a case of higher education institutes in india](#). *Journal of Product & Brand Management*, 22(1):40–51, 2013. doi:10.1108/10610421311298687.
- [4] Thiago de Paulo Faleiros e Alneu de Andrade Lopes. Modelos probabilísticos de tópicos: desvendando o latent dirichlet allocation, 2016.
- [5] Yang Yiming e Liu Xin. A re-examination of text categorization methods. *22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1999.
- [6] Facebook for developers Graph API Overview. [Overview - graph api](#).
- [7] Center for World University Rankings. [Methodology](#), 2017.
- [8] IBM Knowledge Center. [Binning \(grouping\) data values](#), 2017.
- [9] ItBussinessEdge. [What is stemming?](#), 2017.
- [10] L. S. Lai and To. Content analysis of social media: A grounded theory approach. *Journal of Electronic Commerce Research*, 16(2):138–152, 2015.
- [11] Luciana Oliveira and Álvaro Figueira. Benchmarking analysis of social media strategies in the higher education sector. *Procedia Computer Science*, 64:779–786, 2015.
- [12] Luciana Oliveira and Álvaro Figueira. [Social media content analysis in the higher education sector](#). *International Journal of Web Portals*, 7(2):16–32, 2015. doi:10.4018/ijwp.2015040102.
- [13] Luciana Oliveira and Álvaro Figueira. Analyzing social media discourse an approach using semi-supervised learning. *In Proceedings of the 12th International Conference on Web Information Systems and Technologies (WEBIST 2016)*, 2:188–195, 2016.

-
- [14] Baomin Qi and Lindsey Mackie. Utilising social media technology to raise brand awareness in higher education. In *WEBIST (1)*, pages 400–405, 2014.
 - [15] Hugo José Freixo Rodrigues. Ferramenta para text mining em textos completos, 2016.
 - [16] James Sanger Ronen Feldman. *The Text Mining Handbook Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press, 2006.
 - [17] António Paulo Gomes Dos Santos. Classificação multi-etiqueta hierárquica de textos segundo a taxonomia acm, 2008.
 - [18] Julia Silge and David Robinson. *Text Mining with R*. O'REILLY, 2017-04-23.
 - [19] Ah-Hwee Tan. Text mining: The state of the art the chanllenges. *PAKDD 1999 Workshop on Knowledge Disoccovery from Advanced Databases*, 1999.
 - [20] Mathupayas Thongmak. Engaging facebook users in brand pages: Different posts of marketing-mix information. In *International Conference on Business Information Systems*, pages 299–308. Springer, 2015.
 - [21] Luís Torgo. *Data Mining with R Learning with Case Studies*. Chapman & Hall/CRC, 2nd edition, 2011.
 - [22] Luciana Vieira Castilho Weinert. [Ontologias e técnicas de inteligência artificial aplicadas ao diagnóstico em fisioterapia neuropediátrica](#), 2017.

Apêndice A

Análise dos tópicos contidos nas mensagens

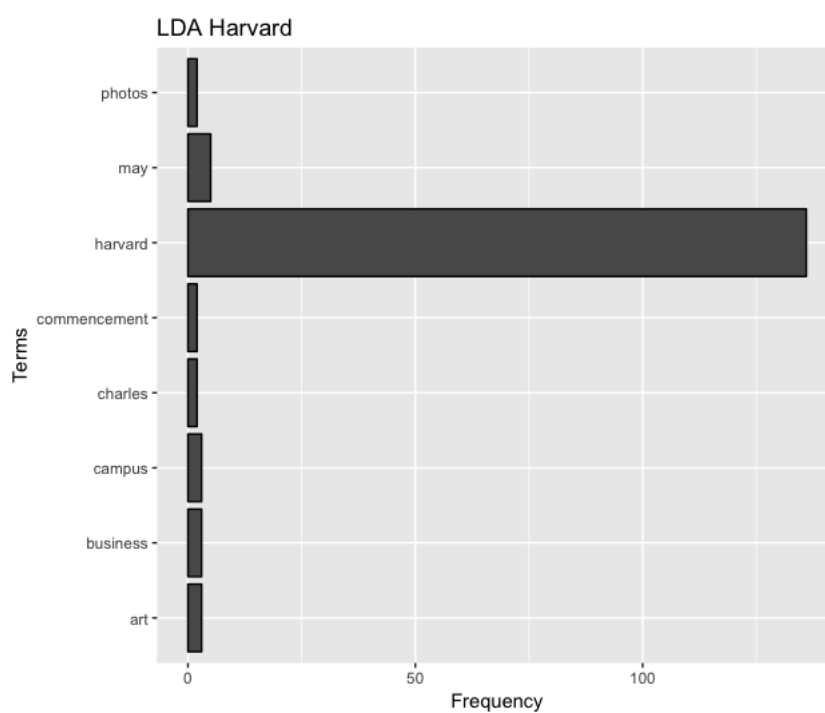
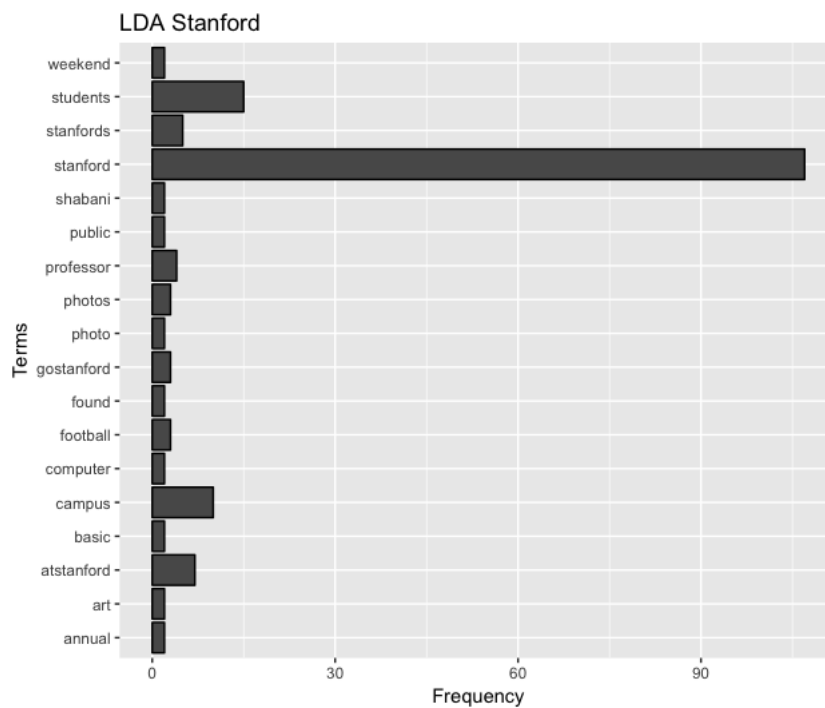
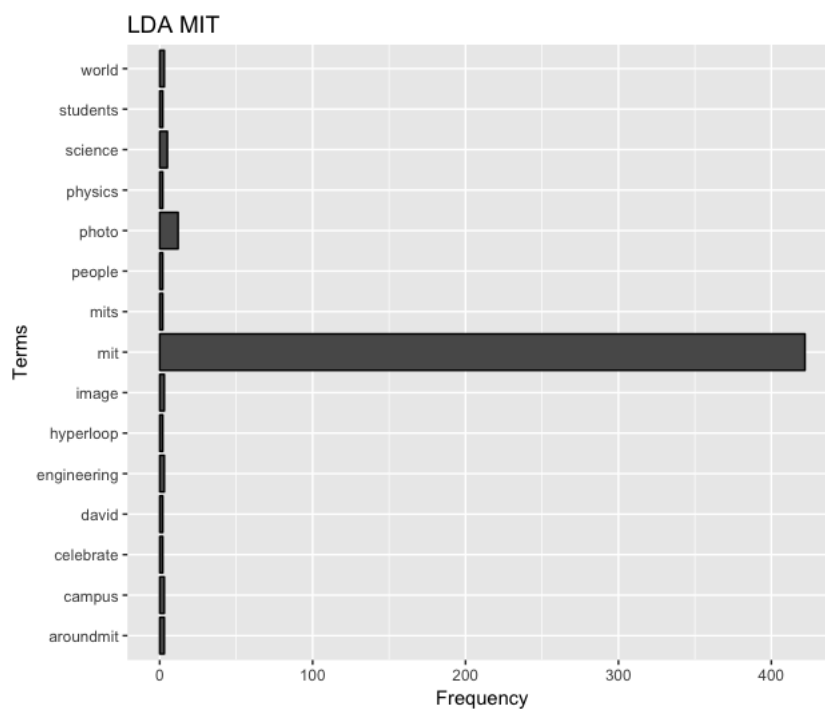
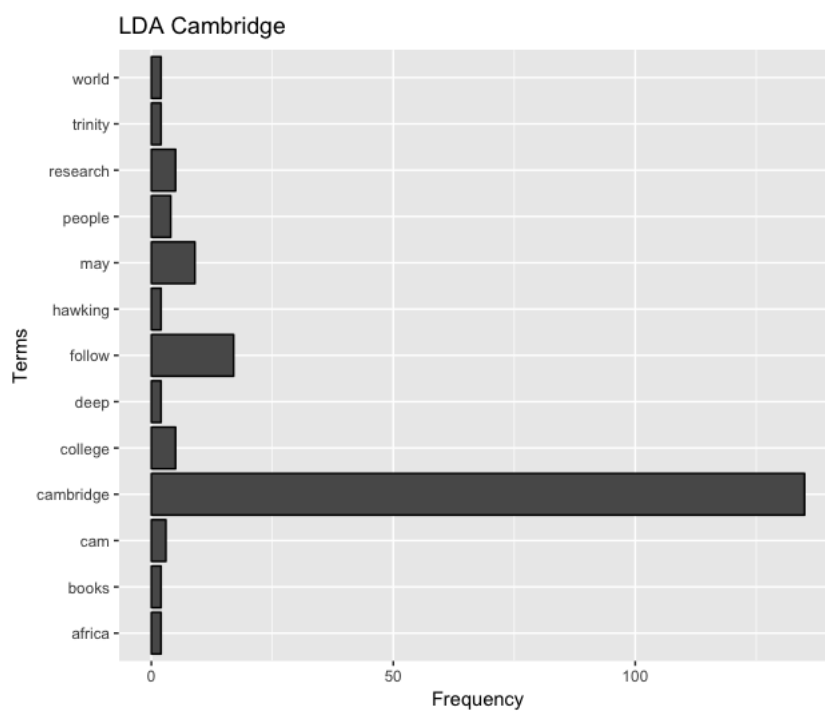
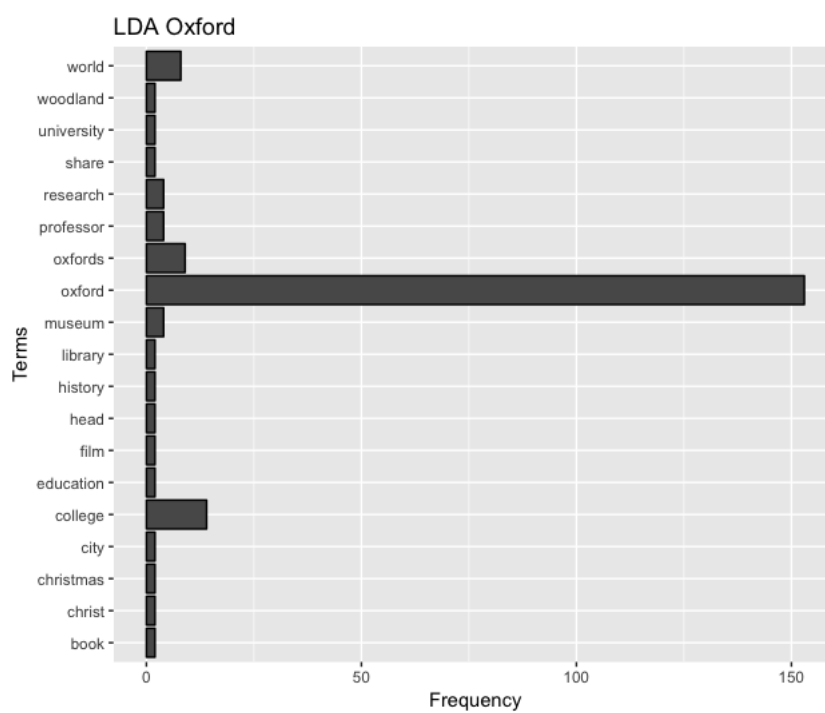
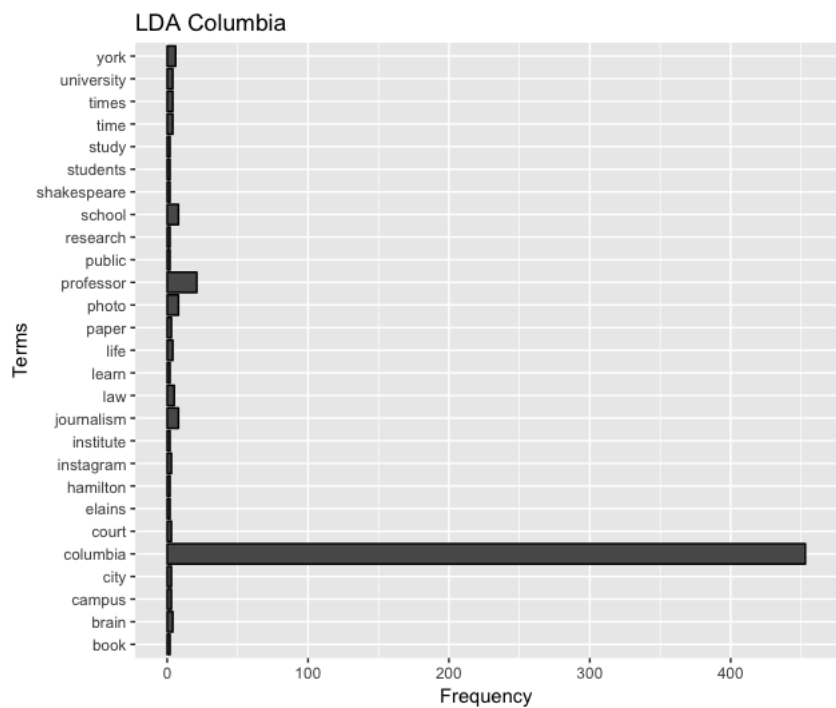
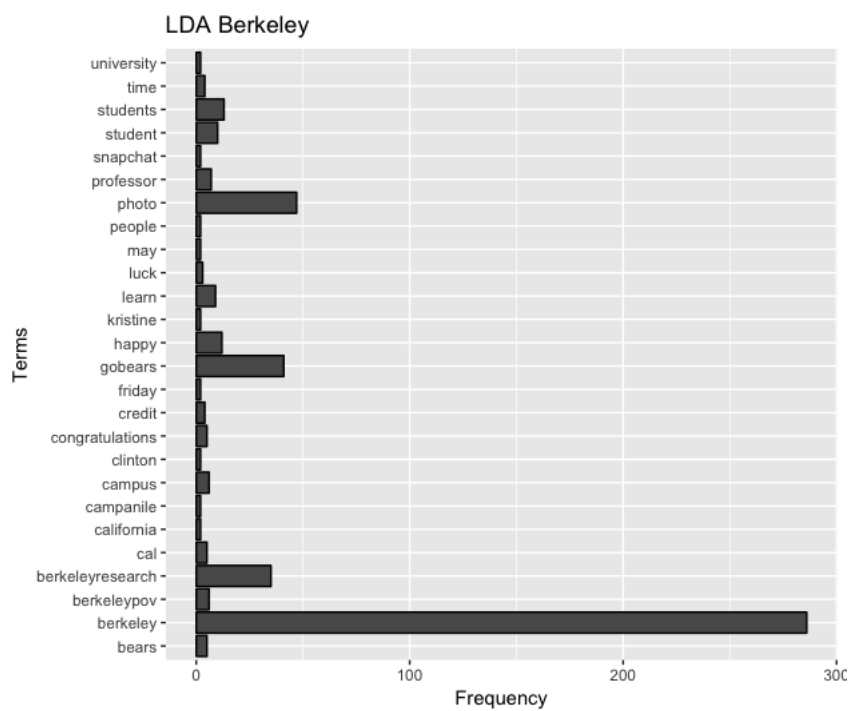
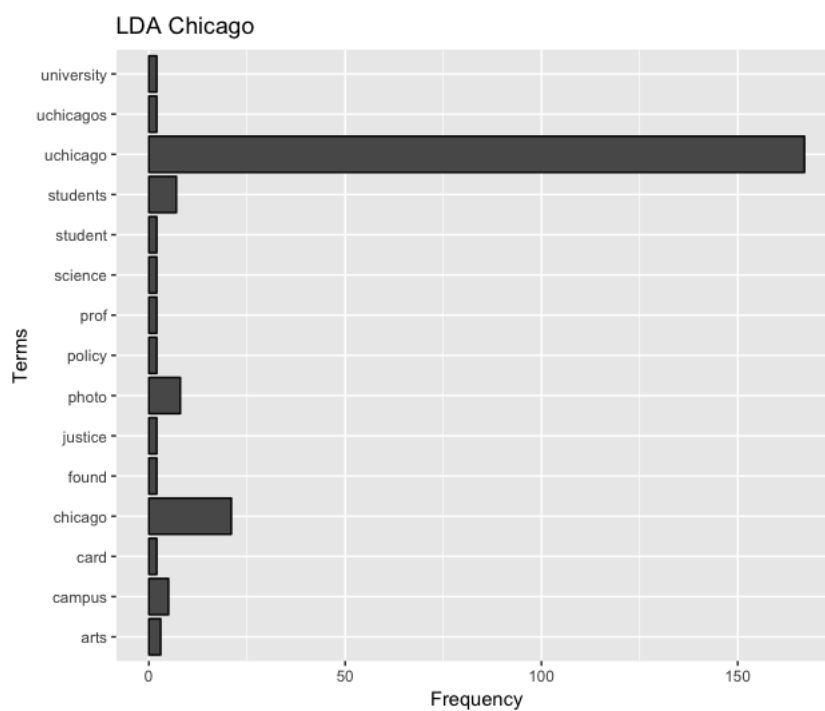
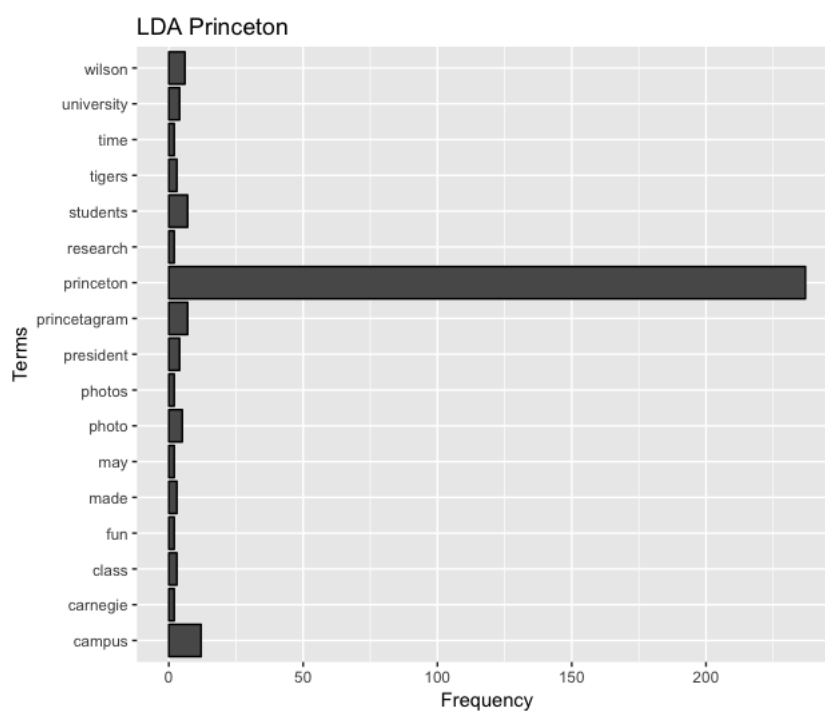


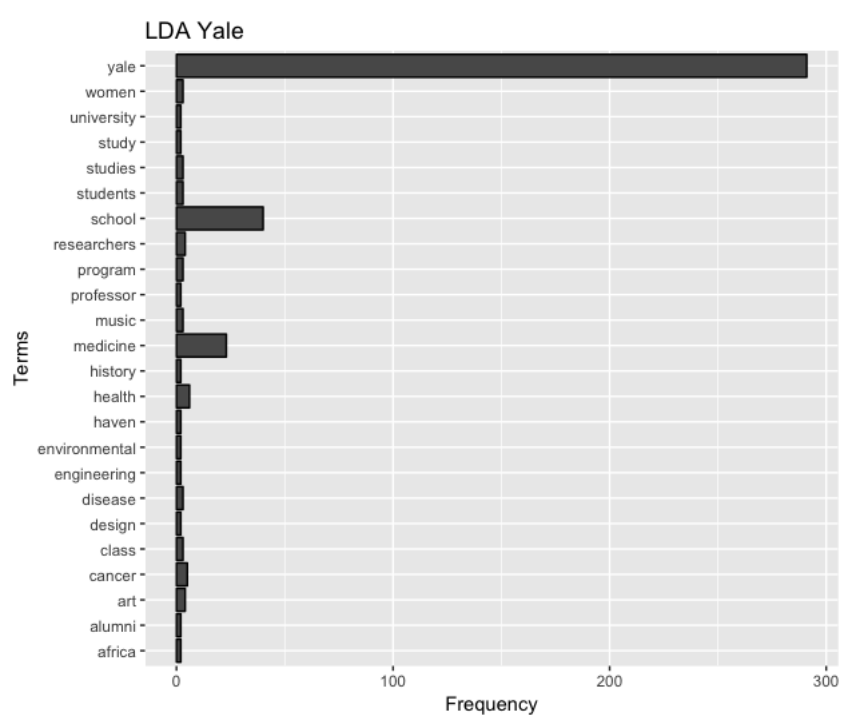
Figura A.1: Tópicos obtidos em *Harvard*

Figura A.2: Tópicos obtidos em *Stanford*Figura A.3: Tópicos obtidos em *MIT*

Figura A.4: Tópicos obtidos em *Cambridge*Figura A.5: Tópicos obtidos em *Oxford*

Figura A.6: Tópicos obtidos em *Columbia*Figura A.7: Tópicos obtidos em *Berkeley*

Figura A.8: Tópicos obtidos em *Chicago*Figura A.9: Tópicos obtidos em *Princeton*

Figura A.10: Tópicos obtidos em *Yale*

Apêndice B

Calendários escolares

Calendário Acadêmico: 2015-16

Anos Calendário Acadêmico: 2015-16, 2014-15⁵⁶, 2013-14⁵⁷

ENCONTRO	EVENTO
18 a 29 de agosto de 2015	LL.M. A orientação do aluno começa
27 de agosto a 1 de setembro de 2015	JD Student Orientation
2 de setembro de 2015	Classes de período de queda 1L começam (as classes seguem uma programação de segunda-feira)
7 de setembro de 2015	Feriado do Dia do Trabalho - Sem aulas; Os escritórios de administração da universidade serão fechados.
8 de setembro de 2015	Classes de período de outono de nível superior começam
12 de outubro de 2015	Dia de Colombo - Férias para todos os alunos; Os escritórios de administração da universidade serão fechados.
13 de outubro de 2015	Sem aulas para todos os alunos
11 de novembro de 2015	Dia do veterano - Todas as aulas se encontram no horário regular; Os escritórios de administração da universidade serão fechados.
25-27 de novembro de 2015	Feriado de Ação de Graças
2 de dezembro de 2015	Fim de 1L do período de outono final
7 de dezembro de 2015	Fim do período de queda do nível superior final
11 a 18 de dezembro de 2015	Período de exame de prazo de queda 1L
14-22 de dezembro de 2015	Período de exame de prazo de queda do nível superior
23 de dezembro de 2015 - 3 de janeiro de 2016	Férias de inverno
4 de janeiro de 2016	Classes de prazo de inverno começam
18 de janeiro de 2016	Dia de Martin Luther King - Férias para todos os alunos; Os escritórios de administração da universidade serão fechados.
20 de janeiro de 2016	Classes de fim de inverno terminam para estudantes de nível superior
21 de janeiro de 2016	Classes de prazo de inverno terminam por 1Ls
22 de janeiro de 2016	Exames de termo de inverno
25 de janeiro de 2016	Todas as Classes de Primavera de Primavera começam
15 de fevereiro de 2016	Dia dos Presidentes - Todas as aulas se encontram no horário regular; Os escritórios de administração da universidade serão fechados.
13 a 20 de março de 2016	Férias de primavera

ENCONTRO	EVENTO
22 de abril de 2016	Terminações de prazo de primavera terminam para todos os alunos
27 de abril a 6 de maio de 2016	Período de exame eletivo de prazo de primavera de nível superior e IL
9 de maio de 2016	Período de exame de lei internacional / comparativa (IL e LLM)
11 a 13 de maio de 2016	IL Cursos Requeridos Período de Exame de Termo de Primavera
26 de maio de 2016	Dia de início

Links

1. <http://hls.harvard.edu/dept/academics/>
2. <http://hls.harvard.edu/dept/academics/academic-calendar/>
3. <http://hls.harvard.edu/dept/academics/degree-programs/>
4. <http://hls.harvard.edu/dept/academics/degree-programs/j-d-program/>
5. <http://hls.harvard.edu/dept/lrw/>
6. <http://hls.harvard.edu/dept/jdadmissions/>
7. <http://hls.harvard.edu/dept/graduate-program/>
8. <http://hls.harvard.edu/dept/academics/degree-programs/special-programs/>
9. <http://hls.harvard.edu/dept/academics/degree-programs/special-programs/joint-jd-master-degree-programs/>
10. <http://hls.harvard.edu/dept/academics/degree-programs/special-programs/joint-jd-master-degree-programs/joint-degree-program-in-law-business/>
11. <http://hls.harvard.edu/dept/academics/degree-programs/special-programs/joint-jd-master-degree-programs/joint-degree-program-in-law-government/>
12. <http://hls.harvard.edu/dept/academics/degree-programs/special-programs/joint-jd-master-degree-programs/joint-degree-in-law-and-public-health/>
13. <http://hls.harvard.edu/dept/academics/degree-programs/special-programs/joint-jd-master-degree-programs/joint-program-in-law-and-urban-planning/>
14. <http://hls.harvard.edu/dept/academics/degree-programs/special-programs/jdphd-program/>
15. <http://hls.harvard.edu/dept/ils/hls-cambridge-joint-degree/>
16. <http://hls.harvard.edu/dept/academics/degree-programs/special-programs/concurrent-degrees/>
17. <http://hls.harvard.edu/dept/ils/semester-abroad/>
18. <http://hls.harvard.edu/dept/ils/winter-term-abroad/>
19. <http://hls.harvard.edu/dept/academics/programs-of-study/>
20. <http://cjp.law.harvard.edu/>
21. <http://hls.harvard.edu/dept/academics/programs-of-study/international-and-comparative-law/>
22. <http://hls.harvard.edu/dept/academics/programs-of-study/law-and-business/>
23. <http://hls.harvard.edu/dept/academics/programs-of-study/law-and-government/>
24. <http://hls.harvard.edu/dept/academics/programs-of-study/law-and-history/>
25. <http://hls.harvard.edu/dept/academics/programs-of-study/law-and-social-change/>
26. <http://hls.harvard.edu/dept/academics/programs-of-study/law-science-and-technology/>
27. <http://hls.harvard.edu/dept/academics/curriculum/>
28. <https://hls.harvard.edu/academics/curriculum/catalog/index.html>
29. <http://hls.harvard.edu/dept/academics/curriculum/course-and-schedule-updates/>
30. <http://hls.harvard.edu/dept/registrar/registration-information/>
31. <http://hls.harvard.edu/dept/registrar/registration-information/1l-registration-information/>
32. http://hls.harvard.edu/?page_id=124074

[Home](#) » [Resources and Help](#) » [Stanford Academic Calendar 2017-18](#) » [Previous Academic Calendars](#)

Academic Calendar

2015-16 Stanford Academic Calendar

Stanford Academic Calendar, 2015-16

- Next year's calendar: [2016-17 Stanford Academic Calendar](#). Note that Summer Quarter 2016 is part of the 2015-16 Academic Calendar.

[Autumn Quarter](#) • [Winter Quarter](#) • [Spring Quarter](#) • [Summer Quarter](#)

How to Feed the Academic Calendar to your Personal Calendar or Receive Text or Email Notifications

The University bill is issued on the 20th of each month. Billed charges are due by the 15th of the following month. For more information, see the [Student Financial Services](#) site or [Stanford ePay](#).

See the [R&DE Student Housing](#) site for additional information on Housing deadlines. Key undergraduate dates are included below.

[Legal-size Printable Calendar](#) • [Letter-size Printable Calendar](#)

Issued April 15, 2015. All dates are subject to change at the discretion of the University.

AUTUMN QUARTER 2015-16

August 1 (Sat) Axxess opens for course enrollment. Stanford Bulletin opens with AY 2015-16 degree requirements

August 24 (Mon) M.D. first-year students, first day of instruction.

August 27 (Thu) M.D. second-year students, first day of instruction.

August 31 (Mon) Law School instruction begins for 1st-year J.D. students. See the full Stanford Law School academic calendar.

September 11 (Fri, 5:00 p.m.) At-status enrollment deadline in order to receive stipend or financial aid refund within the first week of term.

September 11 (Fri) MBA first-year instruction begins.

September 15 (Tue) New undergraduates arrive. Convocation. Undergraduate housing opens for new students; see 2015-16 Student Housing Calendar.

September 17 (Thu) Undergraduate housing opens for returning students; see 2015-16 Student Housing Calendar.

September 21 (Mon) First day of quarter; instruction begins.

September 21 (Mon, 5:00 p.m.) Preliminary Study List deadline. Students must be "at status"; i.e., students must have a study list with sufficient units to meet requirements for their status, whether full-time, 8-9-10 units (graduate students only), or approved Undergraduate Special Registration Status or Graduate Special Registration Status. The late study list fee is \$200.

September 21 (Mon, 5:00 p.m.) Deadline to submit Leave of Absence for full refund (see undergraduate leaves of absence and graduate leaves of absence). See Tuition and Refund Schedule 2015-16 for a full refund schedule.

September 21 (Mon) Law School instruction begins for 2nd/3rd-year J.D. and advanced degree students. See the full Stanford Law School academic calendar.

September 23 (Wed) Yom Kippur (classes held: some students will be observing Yom Kippur and are not expected to attend classes; some faculty will not be holding classes).

September 24 (Thu) Conferral of degrees, Summer Quarter 2014-15.

September 25 (Fri) GSB course add/drop deadline (GSB courses only).

October 9 (Fri, 5:00 p.m.) Final Study List deadline, except GSB. Last day to add or drop a class; last day to adjust units on a variable-unit course. Last day for tuition reassessment for dropped courses or units. Students may withdraw from a course until the Course Withdrawal deadline and a 'W' notation will appear on the transcript.

November 2 (Mon, 5:00 p.m.) Term withdrawal deadline; last day to submit Leave of Absence to withdraw from the University with a partial refund.

November 13 (Fri, 5:00 p.m.) Change of grading basis deadline, except GSB.

November 13 (Fri, 5:00 p.m.) Course withdrawal deadline, except GSB, Law, and M.D.

November 13 (Fri, 5:00 p.m.) Application deadline for Autumn Quarter degree conferral.

November 20 (Fri) Last day of Law classes. See the full Stanford Law School academic calendar.

November 23-27 (Mon-Fri) Thanksgiving Recess (no classes).

November 30-December 6 (Mon-Sun) End-Quarter Period.

December 4 (Fri) Last day of classes (unless class meets on Sat.)

December 4 (Fri) Last opportunity to arrange Incomplete in a course, at last class.

December 4 (Fri, noon) University thesis, D.M.A. final project, or Ph.D. dissertation, last day to submit.

December 4 (Fri, 5:00 p.m.) Late application deadline for Autumn Quarter degree conferral (\$50 fee).

December 4-11 (Fri-Fri) Law School examinations. See the full Stanford Law School academic calendar.

December 7-11 (Mon-Fri) End-Quarter examinations.

December 12 (Sat, noon) Undergraduate housing closes for Winter Break; see 2014-15 Student Housing Calendar.

December 15 (Tue, 11:59 p.m.) Grades due.

January 7 (Thu) Conferral of degrees, Autumn Quarter.

[Top of page](#)

WINTER QUARTER 2015-16

October 25 (Sun) Axess opens for course enrollment.

December 25 (Fri) At-status enrollment deadline in order to receive stipend or financial aid refund within the first week of term.

January 2 (Sat, 8:00 a.m.) Undergraduate housing opens for Winter Quarter; see 2015-16 Student Housing Calendar.

January 4 (Mon) First day of quarter; instruction begins for all students.

January 4 (Mon, 5:00 p.m.) Preliminary Study List deadline. Students must be "at status"; i.e., students must have a study list with sufficient units to meet requirements for their status, whether full-time, 8-9-10 units (graduate students only), or approved Undergraduate Special Registration Status or Graduate Special Registration Status. The late study list fee is \$200.

January 4 (Mon, 5:00 p.m.) Deadline to submit Leave of Absence for full refund (see undergraduate leaves of absence and graduate leaves of absence). See Tuition and Refund Schedule 2015-16 for a full refund schedule.

January 8 (Fri) GSB course add/drop deadline (GSB courses only).

January 18 (Mon) Martin Luther King, Jr., Day (holiday, no classes).

January 22 Fri (5:00 p.m.) Final Study List deadline, except GSB. Final day to add or drop a class; last day to adjust units on a variable-unit course. Last day for tuition reassessment for dropped courses or units. Students may withdraw from a course until the Course Withdrawal deadline and a 'W' notation will appear on the transcript.

February 15 (Mon) Presidents' Day (holiday, no classes; Law does hold classes).

February 17 (Wed, 5:00 p.m.) Term withdrawal deadline; last day to submit Leave of Absence to withdraw from the University with a partial refund.

February 26 (Fri, 5:00 p.m.) Change of grading basis deadline, except GSB.

February 26 (Fri, 5:00 p.m.) Course withdrawal deadline, except GSB, Law, and M.D.

February 26 (Fri, 5:00 p.m.) Application deadline for Winter Quarter degree conferral.

March 7 (Mon) Last day of Law classes. See the full Stanford Law School academic calendar.

March 7-13 (Mon-Sun) End-Quarter Period.

March 11 (Fri) Last day of classes (unless class meets on Sat.)

March 11 (Fri) Last opportunity to arrange Incomplete in a course, at last class.

March 11 (Fri, noon) University thesis, D.M.A. final project, Ph.D. dissertation, last day to submit.

March 11 (Fri, 5:00 p.m.) Late application deadline for Winter Quarter degree conferral (\$50 fee).

March 11-18 (Fri-Fri) Law School examinations. See the full Stanford Law School academic calendar.

March 14-18 (Mon-Fri) End-Quarter examinations.

March 19 (Sat) Undergraduate housing move-out (if departing at end of Winter Quarter); see 2015-16 Student Housing Calendar.

March 22 (Tue, 11:59 p.m.) Grades due.

March 31 (Thu) Conferral of degrees, Winter Quarter.

[Top of page](#)

ACADEMIC CALENDAR 2015 - 2016

June 2015

Su	Mo	Tu	We	Th	Fr	Sa
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30				

8 Monday

Classes of Regular Summer Session begin.

12 Friday

DEGREE APPLICATION DEADLINE for September SB and Advanced Degrees. \$50 Late Fee (\$85 after July 10).

REGISTRATION DEADLINE. Summer Registration for all students **MUST** be submitted by this date. \$50 Late Fee.

15 Monday

Last day to sign up for family health insurance or waive individual coverage for summer, E23-308.

FALL PRE-REGISTRATION DEADLINE. CONTINUING students must initiate online pre-registration by this date. \$50 Late Fee. (\$85 after August 18).

July 2015

Su	Mo	Tu	We	Th	Fr	Sa
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30	31	

3 Friday

Independence Day (Observed) -- Holiday

10 Friday

Last day to submit Advanced Degree Thesis Title. \$85 Late Fee.

August 2015

Su	Mo	Tu	We	Th	Fr	Sa
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

7 Friday

DEADLINE FOR DOCTORAL STUDENTS to submit application, signed by department, to the Office of the Dean for Graduate Education, 3-138, for Fall Term Non-Resident status (\$100 Late Fee).

THESIS DUE for ALL September degree candidates.

14 Friday

LAST DAY TO GO OFF THE SEPTEMBER DEGREE LIST.

Last day of classes for Regular Summer Session.

17, 18 Monday,
Tuesday

Summer Session Final Exam Period.

18 Tuesday

5:00 PM CONTINUING STUDENT FINAL DEADLINE TO PRE-REGISTER ONLINE (\$85 Late Fee).

30 Sunday

Graduate Student Orientation activities begin.

31 Monday

ONLINE REGISTRATION OPENS for all students.

FALL TERM -- Number of Class Days (Wednesday, Sept 9 through Thursday, Dec. 10): 13 Mondays, 12 Tuesdays, 13 Wednesdays, 13 Thursdays, 12 Fridays = 63 days

September 2015

Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30			

- 1 Tuesday English Evaluation Test for new International graduate students, 9:00 am - 12:00 pm.
- 7 Monday Labor Day -- Holiday.
- 8 Tuesday **REGISTRATION DAY** -- Fall Term.
- 9 Wednesday **FIRST DAY OF CLASSES.**
- 11 Friday **DEGREE APPLICATION DEADLINE** for February SB and Advanced Degrees. \$50 Late Fee (\$85 after December 11).
- REGISTRATION DEADLINE.** Registration for all students **MUST** be submitted by this date. \$50 Late Fee.
- 14 Monday First quarter Physical Education classes begin.
- 15 Tuesday Last day to sign up for family health insurance or waive individual coverage for fall, E23-308.
- 16 Wednesday Faculty Officers recommend degrees to Corporation (Degree Award Date).
- 21 - 25 Mon - Fri Career Week
- 25 Friday Fall Career Fair
- MINOR COMPLETION DATE.** Deadline for submission of Minor Completion Form for final-term seniors. \$50 Late Fee.

October 2015

Su	Mo	Tu	We	Th	Fr	Sa
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

- 9 Friday **ADD DATE.** Last day to add subjects to Registration.
- Last day for Graduate students to change a subject to or from P/D/F Grading.
- Last day to change a subject from Listener to Credit.
- Last day to drop half-term subjects offered in first half of term.
- Late fee (\$100) and petition required for students completing registration after this date.
- Deadline for completing cross-registration. \$50 Late Fee for petitions received after this date.
- 12 Monday Columbus Day -- Holiday.
- 13 Tuesday **MONDAY SCHEDULE OF CLASSES TO BE HELD.**
- 23, 24 Friday, Saturday Family Weekend.
- 26 Monday Second quarter Physical Education classes begin.

November 2015

Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30					

- 11 Wednesday Veterans Day -- Holiday.
- 18 Wednesday **DROP DATE.** Last day to cancel subjects from Registration.
- Last day to change a subject from Credit to Listener.
- Last day to add a time-arranged subject that started after beginning of the term.
- Last day to add half-term subjects offered in second half of term.
- 26, 27 Thursday, Friday Thanksgiving Vacation.

December 2015

Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

1 Tuesday

ONLINE PRE-REGISTRATION for Spring Term and IAP begins.

4 Friday

SUBJECTS WITH FINAL EXAM -- no test may be given and no assignment, term paper or oral presentation shall fall due after this date.

SUBJECTS WITH NO FINAL EXAM --

Undergraduate Subjects: no test may be given and at most one assignment may fall due between this date and the end of the last scheduled class period in the subject.

Graduate Subjects: either one in-class test may be given or one assignment may fall due between this date and the end of the last regularly scheduled class in the subject.

10 Thursday

LAST DAY OF CLASSES.

Last day to drop half-term subjects offered in the second half of term.

11 Friday

Last day to submit Advanced Degree Thesis Title. \$85 Late Fee.

14 - 18 Mon - Fri

FINAL EXAM PERIOD.

30 Wednesday

SPRING PRE-REGISTRATION DEADLINE. CONTINUING students must INITIATE online pre-registration by 5:00 pm on this date. \$50 Late Fee (\$85 after January 15).

IAP PRE-REGISTRATION DEADLINE.

Deadline for all students to pre-register for IAP

January 2016

Su	Mo	Tu	We	Th	Fr	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30
31						

4 Monday

First Day of January Independent Activities Period.

DEADLINE FOR DOCTORAL STUDENTS to submit application, signed by department, to the Office of the Dean for Graduate Education, 3-138, for Spring Term Non-Resident status. (\$100 Late Fee). Not needed if Spring Term approved with Fall Term application.

IAP Physical Education classes begin

8 Friday

THESIS DUE for doctoral degrees.

15 Friday

THESIS DUE for degrees other than doctoral.

5:00 PM FINAL DEADLINE FOR CONTINUING STUDENTS TO PRE-REGISTER ONLINE FOR SPRING (\$85 Late Fee).

LAST DAY TO GO OFF THE FEBRUARY DEGREE LIST.

18 Monday

Martin Luther King, Jr. Day -- Holiday.

25 Monday

ONLINE REGISTRATION OPENS for all students.

28 Thursday

English Evaluation Test for new International graduate students, 9:00 am - 12:00 pm.

29 Friday

Last day of January Independent Activities Period.

SPRING TERM -- Number of Class Days (Tuesday, Feb. 2 through Thursday, May 12): 12 Mondays, 12 Tuesdays, 14 Wednesdays, 14 Thursdays, 13 Fridays = 65 days.

February 2016

Su	Mo	Tu	We	Th	Fr	Sa
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29					

1 Monday

REGISTRATION DAY -- Spring Term.

2 Tuesday

FIRST DAY OF CLASSES.

5 Friday

REGISTRATION DEADLINE. Registration for all students MUST be submitted by this date. \$50 Late Fee.

8 Monday

Third quarter Physical Education classes begin.

15 Monday

Presidents Day -- Holiday

Last day to sign up for family health insurance or waive individual coverage for spring, E23-308.

16 Tuesday

MONDAY SCHEDULE OF CLASSES TO BE HELD.

17 Wednesday

Faculty Officers recommend degrees to Corporation (Degree Award Date).

March 2016

Su	Mo	Tu	We	Th	Fr	Sa
		1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31		

4 Friday

ADD DATE. Last day to add subjects to Registration.

Last day for Graduate students to change a subject to or from P/D/F Grading.

Last day to change a subject from Listener to Credit.

Last day to drop half-term subjects offered in first half of term.

Late fee (\$100) and petition required for students completing registration after this date.

Deadline for completing cross-registration. \$50 Late Fee for petitions received after this date.

21 - 25 Mon - Fri

Spring Vacation.

30 Wednesday

Fourth quarter Physical Education classes begin.

April 2016

Su	Mo	Tu	We	Th	Fr	Sa
					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

1 Friday

Last day to submit Advanced Degree Thesis Title. \$85 Late Fee..

18, 19 Monday,
Tuesday

Patriots Day -- Vacation.

21 Thursday

DROP DATE. Last day to cancel subjects from Registration.

Last day to change a subject from Credit to Listener.

Last day to add a time-arranged subject that started after beginning of the term.

Last day to add half-term subjects offered in second half of term.

29 Friday

THESIS DUE for doctoral degrees.

May 2016

Su	Mo	Tu	We	Th	Fr	Sa
1	2	3	4	5	6	7
8	9	10	11	12	13	14
15	16	17	18	19	20	21
22	23	24	25	26	27	28
29	30	31				

2 Monday

ONLINE PRE-REGISTRATION for Fall Term and Summer Session begins.

6 Friday

SUBJECTS WITH FINAL EXAM -- no test may be given and no assignment, term paper or oral presentation shall fall due after this date.

SUBJECTS WITH NO FINAL EXAM --

Undergraduate Subjects: no test may be given and at most one assignment may fall due between this date and the end of the last scheduled class period in the subject.

Graduate Subjects: either one in-class test may be given or one assignment may fall due between this date and the end of the last regularly scheduled class in the subject.

THESIS DUE for degrees other than doctoral.

LAST DAY OF CLASSES.

Last day to drop half-term subjects offered in the second half of term.

FINAL EXAM PERIOD.**LAST DAY TO GO OFF THE JUNE DEGREE LIST.**

Memorial Day -- Holiday.

SUMMER SESSION PRE-REGISTRATION DEADLINE. Deadline for all students to pre-register online for Summer Session. \$50 Late Fee.

12 Thursday

16 - 20 Mon - Fri

20 Friday

30 Monday

31 Tuesday

June 2016

Su	Mo	Tu	We	Th	Fr	Sa
			1	2	3	4
5	6	7	8	9	10	11
12	13	14	15	16	17	18
19	20	21	22	23	24	25
26	27	28	29	30		

2 Thursday

3 Friday

6 Monday

10 Friday

Doctoral Hooding Ceremony.

COMMENCEMENT.

First day of classes for Regular Summer Session.

DEGREE APPLICATION DEADLINE for September SB and Advanced Degrees. \$50 Late Fee (\$85 after July 8).

REGISTRATION DEADLINE. Registration for all students MUST be submitted by this date. \$50 Late Fee.

13 Monday

FALL PRE-REGISTRATION DEADLINE.

CONTINUING students must initiate online pre-registration by this date. \$50 Late Fee. (\$85 after August 16).

DEADLINE FOR CONTINUING STUDENTS to select preferences for Fall CI-H/CI-HW subjects.

JUNE 6 (Mon) - AUG 16 (Tues)SUMMER SESSION (incl. Exam Period). Theses due for all September degree candidates, Friday, August 5.



MATRICULATION, RESIDENCE, ADMISSION TO DEGREES, DISCIPLINE

Year	Full Mich. Term		Full Lent Term		Easter Day	Full Easter Term		General Admission June Th./Fri./Sat.
	begins Oct. Tu.	ends Dec. Fri.	begins Jan. Tu.	ends March Fri.		begins Apr. Tu.	ends June Fri.	
2012–13	2	30 Nov.	15	15	31 Mar.	23	14	27/28/29
2013–14	8	6	14	14	20 Apr.	22	13	26/27/28
2014–15	7	5	13	13	5 Apr.	21	12	25/26/27
2015–16	6	4	12	11	27 Mar.	19	10	23/24/25
2016–17	4	2	17	17	16 Apr.	25	16	29/30/1 July
2017–18	3	1	16	16	1 Apr.	24	15	28/29/30
2018–19	2	30 Nov.	15	15	21 Apr.	23	14	27/28/29
2019–20	8	6	14	13	12 Apr.	21	12	25/26/27
2020–21	6	4	19	19	4 Apr.	27	18	1/2/3 July
2021–22	5	3	18	18	17 Apr.	26	17	30/1/2 July













Academic year 2015-16



















Members of the University can read the full content of each issue online, as webpages or as a PDF, using their single sign-on details (you will be prompted to log in when you view the first of these items on each visit). Non-members of the University may only view the external version of each issue, which omits content covered by data protection legislation, and may not view the PDF versions. All supplements are viewable by everyone.

Trinity term 2016


Date of issue	Web version	PDF version (requires login)	Supplements/Flysheets published with each issue (viewable by all)
21 July 2016	Week 13 No 5141	21 July 2016 - No 5141 (285kb)	Information Security - (1) to No 5140 (78kb)
7 July 2016	Week 11 No 5140	7 July 2016 - No 5140 (180kb)	Anti-Bribery/Anti-Fraud/Public Interest Disclosure - (1) to No 5139 (118kb) Annual Report of the Delegates of the University Press 2015-16
30 June 2016	Week 10 No 5139	30 June 2016 - No 5139 (136kb)	Encaenia 2016 - (1) to No 5138 (161kb)
23 June 2016	Week 9 No 5138	23 June 2016 - No 5138 (133kb)	Bodleian Libraries Annual Report 2014/2015 - (1) to No 5137 (152kb)
16 June 2016	Week 8 No 5137	16 June 2016 - No 5137 (216kb)	
9 June 2016	Week 7 No 5136	9 June 2016 - No 5136 (259kb)	Amended legislative proposal; Resolution concerning Prevent - (1) to No 5135 (131kb)
2 June 2016	Week 6 No 5135	2 June 2016 - No 5135 (182kb)	Professorial Distinction Awards 2016 - No (2) to 5134 (76kb) Flysheet, 2 June 2016 (47kb)
26 May 2016	Week 5 No 5134	26 May 2016 - No 5134 (216kb)	Voting on a resolution: Good Governance in Relation to EJRA - (1) to No 5133 (215kb) Voting on Legislative Proposal: Statute XII - (1) to No 5134 (157kb) Flysheet, 26 May 2016 (67kb)
19 May 2016	Week 4 No 5133	19 May 2016 - No 5133 (203kb)	
12 May	Week 3	12 May 2016 - No	Amendments to legislative proposal: Statutes XII/XI/XIV -

2016	No 5132	5132  (207kb)	(1) to No 5131  (172kb) Flysheet 1, 12 May 2016  (89kb) Flysheet 2, 12 May 2016  (22kb)
5 May 2016	Week 2 No 5131	5 May 2016 - No 5131  (211kb)	
28 April 2016	Week 1 No 5130	28 April 2016 - No 5130  (183kb)	Amendments to Legislative Proposal (Statute XII) - (1) to No 5130  (96kb) Flysheet, 28 April 2016  (89kb)
21 April 2016	Week 0 No 5129	21 April 2016 - No 5129  (251kb)	Lectures and Seminars, Trinity term 2016 - (1) to No 5128  (258kb)

Hilary term 2016

Date of issue	Web version	PDF version (requires login)	Supplements/Flysheets published with each issue (viewable by all)
24 March 2016	Week 10 No 5128	24 March 2016 - No 5128  (121kb)	Oration by the demitting Proctors and Assessor - (1) to No 5127  (82kb) Voting on Legislative Proposal (Statute XII) - (2) to No 5127  (164kb)
17 March 2016	Week 9 No 5127	17 March 2016 - No 5127  (205kb)	
10 March 2016	Week 8 No 5126	10 March 2016 - No 5126  (135kb)	
3 March 2016	Week 7 No 5125	3 March 2016 - No 5125  (164kb)	Student Numbers 2015 - (1) to No 5124  (234kb)
25 February 2016	Week 6 No 5124	25 February 2016 - No 5124  (137kb)	Final Honour Schools 2015 - (1) to No 5123  (230kb)
18 February 2016	Week 5 No 5123	18 February 2016 - No 5123  (168kb)	
11 February 2016	Week 4 No 5122	11 February 2016 - No 5122  (180kb)	
4 February 2016	Week 3 No 5121	4 February 2016 - No 5121  (131kb)	Revised Register of Congregation - (1) to No 5120  (556kb)
28 January 2016	Week 2 No 5120	28 January 2016 - No 5120  (733kb)	
21 January 2016	Week 1 No 5119	21 January 2016 - No 5119  (726kb)	Admission of the Vice-Chancellor - (1) to No 5118  (557kb) University of Oxford Annual Review 2014/15
14 January 2016	Week 0 No 5118	14 January 2016 - No 5118  (798kb)	Lectures and seminars, Hilary term 2016 - (1) to No 5117  (780kb)

Michaelmas term 2015

Date of issue	Web version	PDF version (requires login)	Supplements/Flysheets published with each issue (viewable by all)
10 December 2015	Week 9 No 5117	10 December 2015 - No 5117  (847kb)	

3 December 2015	Week 8 No 5116	3 December 2015 - No 5116  (679kb)	
26 November 2015	Week 7 No 5115	26 November 2015 - No 5115  (669kb)	
19 November 2015	Week 6 No 5114	19 November 2015 - No 5114  (706kb)	Third consultation on possible revision to Statute XII - (1) to No 5113  (691kb)
12 November 2015	Week 5 No 5113	12 November 2015 - No 5113  (736kb)	
5 November 2015	Week 4 No 5112	5 November 2015 - No 5112  (115kb)	
29 October 2015	Week 3 No 5111	29 October 2015 - No 5111  (773kb)	
22 October 2015	Week 2 No 5110	22 October 2015 - No 5110  (699kb)	Recognition of Distinction 2016: Call for applications - (1) to No 5109  (571kb)
15 October 2015	Week 1 No 5109	15 October 2015 - No 5109  (706kb)	Oration by the Vice-Chancellor - (1) to No 5108  (560kb)
8 October 2015	Week 0 No 5108	8 October 2015 - No 5108  (718kb)	Lectures and seminars, Michaelmas term 2015 - (1) to No 5107  (805kb)
1 October 2015	Week 00 No 5107	1 October 2015 - No 5107  (745kb)	
24 September 2015	Week 000 No 5106	24 September 2015 - No 5106  (629kb)	University of Oxford Safeguarding Code of Practice - (1) to No 5106  (547kb)

REGISTRATION DATES

2015 - 2021

REGISTRATION DATES 2015

ARCHITECTURE, PLANNING & PRESERVATION

Spring 2015

Registration Dates

Monday, November 24 - Wednesday, November 26

Tuesday, January 13 - Friday, January 16

Change of Program

Tuesday, January 20 - Friday, January 23

Monday, January 26 - Friday, January 30

Fall 2015

Registration Dates

New Incoming Architecture, Planning & Preservation Students

Tuesday, September 1 - Thursday, September 3

All Returning Architecture, Planning & Preservation Students

Monday, April 13 - Friday, April 17

Tuesday, September 1 - Thursday, September 3

Change of Program

Tuesday, September 8 - Friday, September 11

Monday, September 14 - Friday, September 18

BUSINESS

Spring 2015

Consult Student Affairs Office

Fall 2015

Consult Student Affairs Office

COLUMBIA COLLEGE

Spring 2015

Registration Dates

Rising Seniors

Monday, November 17

Rising Juniors

Tuesday, November 18

Rising Sophomores



Wednesday, November 19

All Returning Columbia College Students

Thursday, November 20 – Friday, November 21

Monday, December 8 – Friday, December 12

Tuesday, January 6 – Friday, January 9

Monday, January 12 – Friday, January 16

Transfers

Tuesday, January 6 – Friday, January 9

Monday, January 12 – Friday, January 16

Change of Program

Tuesday, January 20 – Friday, January 23

Monday, January 26 – Friday, January 30

Fall 2015

Registration Dates

Rising Seniors

Monday, April 13

Rising Juniors

Tuesday, April 14

Rising Sophomores

Wednesday, April 15

All Returning Columbia College Students

Wednesday, April 15 – Friday, April 17

Summer Registration for Fall Term

Monday, June 8 – Friday, June 12

Monday, June 15 – Friday, June 19

Monday, July 27 – Friday, July 31

Monday, August 3 – Friday, August 7 Tuesday,

August 25 – Thursday, August 27 Saturday,

September 5

Transfer Students

Tuesday, August 25 – Thursday, August 27

Saturday, September 5

First Year Students

Thursday, September 3 – Saturday, September 5

Change of Program

Tuesday, September 8 – Friday, September 11

Monday, September 14 – Friday, September 18

PROFESSIONAL STUDIES

Spring 2015

Registration Dates

New Incoming Professional Studies and Auditing Students

Monday, January 12 – Friday, January 16

All Returning Auditing Students



Monday, January 5 – Friday, January 9

Monday, January 12 – Friday, January 16

All Returning Professional Studies Students

Monday, November 17 – Friday, November 21

Monday, November 24 – Wednesday, November 26

Monday, December 1 – Friday, December 5

Monday, December 8 – Friday, December 12

Monday, December 15 – Friday, December 19

Monday, December 22 – Tuesday, December 23

Friday, December 26

Monday, December 29 – Tuesday, December 30

Friday, January 2

Monday, January 5 – Friday, January 9

Monday, January 12 – Friday, January 16

Change of Program

Tuesday, January 20 – Friday, January 23

Monday, January 26 – Friday, January 30

Fall 2015

Registration Dates

New Incoming Executive Trio and Returning Auditing Students

Monday, August 17 – Friday, August 21

Monday, August 24 – Friday, August 28

Monday, August 31 – Friday, September 4

New Incoming Auditing Students

Monday, August 24 – Friday, August 28

Monday, August 31 – Friday, September 4

New Incoming Professional Studies Students

Monday, August 31 – Friday, September 4

All Returning Professional Studies Students

Monday, April 13 – Friday, April 17

Monday, April 20 – Friday, April 24

Monday, April 27 – Friday, May 1

Monday, May 4 – Friday, May 8

Monday, May 11 – Friday, May 15

Monday, May 18 – Friday, May 22

Tuesday, May 26 – Friday, May 29

Monday, June 1 – Friday, June 5

Monday, June 8 – Friday, June 12

Monday, June 15 – Friday, June 19

Monday, June 22 – Friday, June 26

Monday, June 29 – Friday, July 3

Monday, July 6 – Friday, July 10

Monday, July 13 – Friday, July 17

Monday, July 20 – Friday, July 24

Monday, July 27 – Friday, July 31

Monday, August 3 – Friday, August 7

Monday, August 10 – Friday, August 14



Monday, August 17 – Friday, August 21
Monday, August 24 – Friday, August 28
Monday, August 31 – Friday, September 4

Change of Program

Tuesday, September 8 – Friday, September 11
Monday, September 14 – Friday, September 18

ENGINEERING & APPLIED SCIENCE (GRADUATE)

Spring 2015

Registration Dates

Monday, November 24 – Wednesday, November 26
Tuesday, January 13 – Friday, January 16

Change of Program

Tuesday, January 20 – Friday, January 23
Monday, January 26 – Friday, January 30

Fall 2015

Registration Dates

EMFINE, EMMSNE, Second Year EMINEN, and Second Year EMOPRE Students

Monday, April 13
Wednesday, April 15 – Friday, April 17
Monday, July 27 – Friday, July 31
Monday, August 3 – Friday, August 7
Tuesday, September 1 – Thursday, September 3

All Other Graduate Engineering & Applied Science Students

Tuesday, April 14
Wednesday, April 15 – Friday, April 17
Monday, July 27 – Friday, July 31
Monday, August 3 – Friday, August 7
Tuesday, September 1 – Thursday, September 3

Change of Program

Tuesday, September 8 – Friday, September 11
Monday, September 14 – Friday, September 18

ENGINEERING & APPLIED SCIENCE (UNDERGRADUATE)

Spring 2015

Registration Dates

Rising Seniors

Monday, November 17

Rising Juniors

Tuesday, November 18

Rising Sophomores

Wednesday, November 19

All Returning Undergraduate SEAS Students



Thursday, November 20 – Friday, November 21

Monday, December 8 – Friday, December 12

Tuesday, January 6 – Friday, January 9

Monday, January 12 – Friday, January 16

Transfers

Tuesday, January 6 – Friday, January 9

Monday, January 12 – Friday, January 16

Change of Program

Tuesday, January 20 – Friday, January 23

Monday, January 26 – Friday, January 30

Fall 2015

Registration Dates

Rising Seniors

Monday, April 13

Rising Juniors

Tuesday, April 14

Rising Sophomores

Wednesday, April 15

All Returning Undergraduate SEAS Students

Wednesday, April 15 – Friday, April 17

Summer Registration for Fall Term

Monday, June 8 – Friday, June 12

Monday, June 15 – Friday, June 19

Monday, July 27 – Friday, July 31

Monday, August 3 – Friday, August 7 Tuesday,

August 25 – Thursday, August 27 Saturday,

September 5

Transfers

Tuesday, August 25 – Thursday, August 27

Saturday, September 5

First Year Students

Thursday, September 3 – Saturday, September 5

Change of Program

Tuesday, September 8 – Friday, September 11

Monday, September 14 – Friday, September 18

GENERAL STUDIES

Spring 2015

Registration Dates

Monday, November 17 – Friday, November 21

Monday, December 8 – Friday, December 12

Tuesday, January 6 – Friday, January 9

Monday, January 12 – Friday, January 16

Change of Program

Tuesday, January 20 – Friday, January 23

Monday, January 26 – Friday, January 30

Fall 2015

Registration Dates

Monday, April 13 – Friday, April 17

Summer Registration for Fall Term

Monday, June 8 – Friday, June 12

Monday, June 15 – Friday, June 19

Monday, July 27 – Friday, July 31

Monday, August 3 – Friday, August 7

Tuesday, August 25 – Thursday, August 27

Tuesday, September 1 – Thursday, September 3

Change of Program

Tuesday, September 8 - Friday, September 11

Monday, September 14 - Friday, September 18

GRADUATE SCHOOL OF ARTS & SCIENCES

Spring 2015

Registration Dates

Monday, November 24 – Wednesday, November 26

Monday, December 8 – Friday, December 12

Monday, January 5 – Friday, January 9

Monday, January 12 – Friday, January 16

Change of Program

Tuesday, January 20 – Friday, January 23

Monday, January 26 – Friday, January 30

Fall 2015

Registration Dates

Monday, April 13 - Friday, April 17

Summer Registration for Fall Term

Monday, June 8 – Friday, June 12

Monday, June 15 – Friday, June 19

Monday, July 27 – Friday, July 31

Monday, August 3 – Friday, August 7

Tuesday, August 25 – Thursday, August 27

Tuesday, September 1 – Thursday, September 3

Change of Program

Tuesday, September 8 - Friday, September 11

Monday, September 14 - Friday, September 18

Journalism

Spring 2015

Consult Student Affairs Office

Fall 2015

Consult Student Affairs Office

SCHOOL OF INTERNATIONAL AND PUBLIC AFFAIRS

Spring 2015

Registration Dates

New Incoming School of International & Public Affairs Students

Thursday, January 15 – Friday, January 16

All Returning PEPM, EMPA, & ENVP Students

Monday, November 17 – Friday, November 20

Monday, November 24 – Wednesday, November 26

Tuesday, January 13 – Friday, January 16

All Returning School of International & Public Affairs Students

Monday, November 17 – Friday, November 20

Monday, November 24 – Wednesday, November 26

Tuesday, January 13 – Wednesday, January 14

Change of Program

Tuesday, January 20 – Friday, January 23

Monday, January 26 – Friday, January 30

Fall 2015

Registration Dates

New Incoming School of International & Public Affairs Students

Tuesday, September 1 – Friday, September 4

All Returning EMPA Students

Tuesday, April 14 – Friday, April 17

Wednesday, September 2 – Thursday, September 3

All Returning School of International & Public Affairs Students

Tuesday, April 14 – Friday, April 17

Monday, August 24 – Friday, August 28

Change of Program

Tuesday, September 8 – Friday, September 11

Monday, September 14 – Friday, September 18

SCHOOL OF THE ARTS

Spring 2015

Registration Dates

Monday, November 24 – Wednesday, November 26

Tuesday, January 13 – Friday, January 16

Change of Program

Tuesday, January 20 – Friday, January 23

Monday, January 26 – Friday, January 30



Fall 2015

Registration Dates

Monday, April 13 – Friday, April 17

Tuesday, September 1 – Thursday, September 3

Change of Program

Tuesday, September 8 – Friday, September 11

Monday, September 14 – Friday, September 18

SOCIAL WORK

Spring 2015

Registration Dates

Consult Student Affairs Office

Change of Program

Tuesday, January 20 – Friday, January 23

Monday, January 26 – Friday, January 30

Fall 2015

Registration Dates

Consult Student Affairs Office

Change of Program

Tuesday, September 8 – Friday, September 11

Monday, September 14 – Friday, September 18

COLUMBIA LAW SCHOOL

Academic Calendar 2015–2016

Autumn Term 2015

August	
17 and 18	1 st year JD Registration & Orientation
18	LLM Registration & Orientation
19-September 4	1 st year JD classes in <i>Legal Methods (LM)</i> and in <i>Legal Practice Workshop (LPW)</i> . <i>LPW</i> continues throughout the fall term). <i>Legal Methods</i> final exam on Sept. 25
19-September 4	Classes in LLM <i>Introduction to American Law (IAL)</i> and in <i>Legal Research & Writing (LRW)</i> . LLM <i>LRW</i> continues throughout the fall term). <i>IAL</i> final exam on Sept. 25
31	Start of Change of Program Period (Add/Drop) - only voluntary changes until Sept. 8
September	
1	October 2015 Degree Candidates: Deadline to submit outstanding 2014-15 written work granted an extension
4	End of classes in 1L <i>Legal Methods</i> and LLM <i>Introduction to American Law (IAL)</i>
7	Labor Day - No classes
8 (Tuesday)	First day of classes in term-long Law courses. First day of classes at other divisions of the University.
8-18	Law School Change of Program Period (Add/Drop) continues; action required on wait-list notifications
18	End of Law School Change of Program Period; after this date changes only with permission of the Rules Committee, upon petition
18	3L's: Deadline to register for Major Writing Credit
24	<i>Legal Methods</i> Study Day – No 1L classes
25	Final exams for 1L <i>Legal Methods</i> and LLM <i>Introduction to American Law</i>
October	
1	October 2015 LLM Degree Candidates: Last day to file final copy of LLM Essay with Graduate Legal Studies
15	Final date to drop a Fall 2015 Law course (the notation of W will be recorded for these late drops)
15	LLM Writing Project registration deadline for fall term writing projects
15	Final date to submit outstanding 2014-15 written work that was granted an extension (Grades due by Nov. 15)
21	Conferral of October degrees
November	
26-27	Thanksgiving Holiday - No Classes
December	
1	May 2016 Degree Candidates: Degree applications due
4 (Friday)	Last day of 1L classes
5-10	1L study period (1L Exams Dec. 11-22)
11-22	1L exams
11 (Friday)	Last day of upper-year Law classes
12-13	Upper-year study days (Upper-year Exams Dec. 14-22)
14	Last day of classes at other divisions of the University
14-22	Upper-year exams
22	Fall 2015 term ends; due date for submission of Fall 2015 written work
31	February 2016 Degree Candidates: Deadline to submit outstanding written work; last day to file final copy of LLM Essay with Graduate Legal Studies

Spring Term 2016

January	
11	Start of Change of Program Period (Add/Drop) - only voluntary changes until January 19
15	Fall 2015 grades due
18	Martin Luther King, Jr. Birthday Observed - No Classes
19 (Tuesday)	First day of all Law School classes. First day of classes at other divisions of the University
19-26	Law School Change of Program Period (Add/Drop) continues; action required on wait-list notification
26	End of Law School Change of Program Period; after this date changes only with permission of the Rules Committee, upon petition
26	3L's: Deadline to register for Minor Writing Credit
February	
1	3L's: Deadline to submit first draft of Major Writing Credit paper
1	LLM Writing Project registration deadline for spring term writing projects
10	Conferral of February 2016 degrees
15	Final date to drop a Spring 2016 Law course (the notation of W will be recorded for these late drops)
March	
14-18	Spring Recess - No Classes
April	
1	3L's: Deadline to submit final draft of Major Writing Credit paper
1	All students: Deadline to submit all outstanding Fall 2015 written work that was granted an extension
27 (Wednesday)	Last day of Law School classes
27	May 2016 Degree Candidates (JD & LLM): Deadline to submit all outstanding Spring 2016 written work
27	May 2016 LLM Degree Candidates: Last day to file final copy of LLM Essay with Graduate Legal Studies
28-May 1	Upper-year study period (Exams May 2-13)
28-May 3	1L study period (Exams May 4-13)
May	
2	Last day of classes at other divisions of the University
2-13	Upper-year exams
4-13	1L exams
13	Last day of exams for all Law courses
13	Continuing students: Due date for submission of Spring 2016 written work
13	Grades due for May 2017 degree candidates
13	Spring 2016 term ends
18 (Wednesday)	University Commencement & Degree Conferral
19 (Thursday)	Law School Graduation Ceremony
June	
15	Spring 2016 grades due for continuing students

2015-16 BERKELEY ACADEMIC CALENDAR

2015 Fall Semester

Fall Semester Begins	Wednesday, August 19, 2015
Convocation	Monday, August 24, 2015
Instruction Begins	Wednesday, August 26, 2015
Academic and Administrative Holiday	Monday, September 7, 2015
Academic and Administrative Holiday	Wednesday, November 11, 2015
Non-Instructional Day	Wednesday, November 25, 2015
Academic and Administrative Holiday	Thursday, November 26 & Friday, November 27, 2015
Formal Classes End	Friday, December 4, 2015
Reading/Review/Recitation Week	Monday, December 7-Friday, December 11, 2015
Last Day of Instruction	Friday, December 11, 2015
Final Examinations	Monday, December 14-Friday, December 18, 2015
Fall Semester Ends	Friday, December 18, 2015
Winter Commencement	Sunday, December 20, 2015
Academic and Administrative Holiday	Thursday, December 24 & Friday, December 25, 2015
Academic and Administrative Holiday	Thursday, December 31, 2015 & Friday, January 1, 2016

2016 Spring Semester

Spring Semester Begins	Tuesday, January 12, 2016
Academic and Administrative Holiday	Monday, January 18, 2016
Instruction Begins	Tuesday, January 19, 2016
Academic and Administrative Holiday	Monday, February 15, 2016
Spring Recess	Monday, March 21-Friday, March 25, 2016
Academic and Administrative Holiday	Friday, March 25, 2016
Cal Day	Saturday, April 16, 2016
Formal Classes End	Friday, April 29, 2016
Reading/Review/Recitation Week	Monday, May 2-Friday, May 6, 2016
Last Day of Instruction	Friday, May 6, 2016
Final Examinations	Monday, May 9-Friday, May 13, 2016
Spring Semester Ends	Friday, May 13, 2016
Commencement	Saturday, May 14, 2016
Academic and Administrative Holiday	Monday, May 30, 2016

2016 Summer Sessions

First Six-Week Session Begins	Monday, May 23, 2016
Academic and Administrative Holiday	Monday, May 30, 2016
Ten-Week Session Begins	Monday, June 6, 2016
Eight-Week Session Begins	Monday, June 20, 2016
First Six-Week Session Ends	Friday, July 1, 2016
Academic and Administrative Holiday	Monday, July 4, 2016
Second Six-Week Session Begins	Tuesday, July 5, 2016
Three-Week Session Begins	Monday, July 25, 2016
Second Six-Week Session Ends	Friday, August 12, 2016
Ten-Week Session Ends	Friday, August 12, 2016
Eight-Week Session Ends	Friday, August 12, 2016
Three-Week Session Ends	Friday, August 12, 2016



THE UNIVERSITY OF CHICAGO

2015–2016 Academic Year

Summer 2015

Date	Event/Deadline
Monday, June 22	Summer Quarter Begins
Friday, July 3	Independence Day Holiday
Friday, August 28	Convocation
Saturday, August 29	Summer Quarter Ends
Friday, September 4	Pritzker School of Medicine Quarter Ends

Autumn 2015

Date	Event/Deadline
Sunday–Saturday, September 20–26	College Orientation Week
Wednesday, September 23	Incoming Student Registration (College)
Monday, September 28	Autumn Quarter Begins
Thursday–Friday, November 26–27	Thanksgiving Break
Thursday–Friday, December 3–4	College Reading Period
Friday, December 11	Convocation
Saturday, December 12	Autumn Quarter Ends
Sunday–Saturday, September 18–24	College Orientation Week
Wednesday, September 21	Incoming Student Registration (College)

Winter 2016

Date	Event/Deadline
Monday, January 4	Winter Quarter Begins
Monday, January 18	Martin Luther King, Jr. Day
Friday, February 12	College Break
Thursday–Friday, March 10–11	College Reading Period
Friday, March 18	Convocation
Saturday, March 19	Winter Quarter Ends

Spring 2016

Date	Event/Deadline
Monday, March 28	Spring Quarter Begins
Monday, May 30	Memorial Day
Thursday–Friday, June 2–3	College Reading Period
Saturday, June 11	Convocation
Saturday, June 11	Spring Quarter Ends

ACADEMIC CALENDAR 2015-2016			
Princeton University			
Office of the Registrar			
FALL TERM			
September	9	Wednesday	Graduate/Undergraduate Academic Year Sign-In begins at 7:00 a.m.
	13	Sunday	Opening Exercises, University Chapel
	16	Wednesday	Classes begin at 8:00 a.m.
	29	Tuesday	Undergraduate deadline to add or drop courses without fee
October	5	Monday	First day of Ph.D. general examinations
	25	Sunday	Last day of Ph.D. general examinations
	26	Monday	First day of midterm test week
	30	Friday	Last day of midterm test week
	31	Saturday	Fall recess begins
November	8	Sunday	Fall recess ends
	9	Monday	Classes resume
	9	Monday	Undergraduate selection of P/D/F option begins
	9	Monday	Faculty deadline for reporting midterm grades
	24	Tuesday	Thanksgiving recess begins after last class
	29	Sunday	Thanksgiving recess ends
	30	Monday	Classes resume
	30	Monday	Undergraduate deadline last day to drop fall term courses or select P/D/F option
December	8	Tuesday	Graduate students begin spring term course selection
	9-11	Wed-Fri	Senior spring course selection period
	11-15	Fri-Tues	Junior spring course selection period
	14-16	Mon-Wed	Sophomore spring course selection period
	16-18	Wed-Fri	Freshman spring course selection period
	18	Friday	Graduate student deadline for fall term course changes
	18	Friday	Winter recess begins after last class
January	3	Sunday	Winter recess ends
	4	Monday	First day of Ph.D. general examinations
	4	Monday	Reading Period begins
	5	Tuesday	Undergraduate deadline for submitting junior independent work
	12	Tuesday	Reading Period ends
	12	Tuesday	Dean's Date; deadline for student submission of written work
	13	Wednesday	Fall Term examinations begin
	18	Monday	Deadline for submission of take-home exercise
	19	Tuesday	Faculty deadline for submission of grades for undergraduate courses with no scheduled final examination or take-home exercise
	20	Wednesday	Faculty deadline for submission of grades for junior independent work
	20	Wednesday	Faculty deadline for submission of grades for undergraduate courses with take-home exercise (fewer than 20 students)
	22	Friday	Faculty deadline for submission of grades for undergraduate courses with take-home exercise (20 or more students)
	22	Friday	Faculty deadline for submission of graduate course grades
	23	Saturday	Fall Term examinations end
	24	Sunday	Last day of Ph.D. general examinations

ACADEMIC CALENDAR 2015-2016			
Princeton University			
Office of the Registrar			
SPRING TERM			
February	1	Monday	Classes begin 8:00 a.m.
	12	Friday	Undergraduate deadline to add or drop courses without a fee
March	7	Monday	First day of midterm test week
	11	Friday	Last day of midterm test week
	12	Saturday	Spring recess begins
	20	Sunday	Spring recess ends
	21	Monday	Classes resume
	21	Monday	Undergraduate selection of P/D/F option begins
	21	Monday	Faculty deadline for reporting midterm grades
April	8	Friday	Undergraduates deadline to drop spring term courses or select P/D/F option
	18	Monday	First day of Ph.D. General examinations
	11-19	Mon-Tue	AB Sophomore Concentration Declaration period
	18-20	Mon-Wed	Junior fall 2016 course selection period
	20-22	Wed-Fri	Sophomore fall 2016 course selection period
	25-27	Mon-Wed	Freshman fall 2016 course selection period
	29	Friday	Graduate student deadline for spring term course changes
May	2	Monday	Reading Period begins
	2	Monday	Deadline for submitting senior thesis
	3	Tuesday	Deadline for submitting junior independent work
	10	Tuesday	Reading Period ends
	10	Tuesday	Dean's Date: deadline for student submission of written work
	11	Wednesday	Spring term examinations begin
	11	Wednesday	Senior Departmental/Comprehensive examinations begin
	12	Thursday	Senior Departmental/Comprehensive examinations end
	16	Monday	Deadline for submission of take-home exercises
	17	Tuesday	Faculty deadline for submission of grades for undergraduate courses with no scheduled final examination or take-home exercise
	18	Wednesday	Faculty deadline for submission of grades for undergraduate courses with take-home exercise (fewer than 20 students)
	19	Thursday	Faculty deadline for submission of senior thesis and senior comprehensive grades
	20	Friday	Faculty deadline for submission of grades for undergraduate courses with take-home exercise (20 or more students)
	20	Friday	Faculty deadline for submission of graduate course grades
	21	Saturday	Spring Term examinations end
	21	Saturday	Last day of Ph.D. general examinations
	24	Tuesday	Faculty deadline for submission of grades for junior independent work
	25	Wednesday	Honors due to Registrar by Noon
	29	Sunday	Baccalaureate
	30	Monday	Class Day
	30	Monday	Graduate School Hooding Ceremony
	31	Tuesday	Commencement Day, Class of 2016

Revised 10/1/2015

**YALE LAW SCHOOL
CALENDAR FOR ACADEMIC YEAR 2015-2016**

2015

FALL TERM

24	Aug.	Mon.	Orientation for all graduate students.
26 - 28	Aug.	Wed. – Fri.	Orientation for all new JD students. Registration for all new students.
31	Aug.	Mon.	Registration for all returning students. First-term small groups meet.
1	Sept.	Tues.	<u>Fall term begins, 8:10 a.m.</u> Classes begin, 8:10 a.m. Add/drop period begins, 8:30 a.m.
4	Sept.	Fri.	Only Monday classes meet.
7	Sept.	Mon.	Labor Day. No classes will meet.
11	Sept.	Fri.	Add/drop period closes, 4:30 p.m.
8	Oct.	Thurs.	Fall recess begins, 9 p.m.
9 - 12	Oct.	Fri. - Mon.	Fall recess; no classes will meet.
13	Oct.	Tues.	Classes resume, 8:10 a.m.
25	Nov.	Wed.	Only Friday classes will meet. Thanksgiving recess begins, 6 p.m.
30	Nov.	Mon.	Classes resume, 8:10 a.m.
1	Dec.	Tues.	Only Friday classes will meet.
7	Dec.	Mon.	Classes end, 9 p.m.
8- 13	Dec.	Tues. - Sun.	Reading period.
14	Dec.	Mon.	Examination period begins, 8:30 a.m.
22	Dec.	Tues. .	Examinations end, 5 p.m.; <u>fall term ends</u> , 5 p.m.; winter recess begins.

2016

SPRING TERM

18	Jan.	Mon.	MLK Holiday. No classes will meet.
19	Jan.	Tues.	<u>Spring term begins, 8:10 a.m.</u> Registration for all students, 8:30 a.m. – 5 p.m. Classes begin, 8:10 a.m.
20	Jan.	Wed.	Add/drop period begins, 8:30 a.m.
29	Jan.	Fri.	Add/drop period ends, 4:30 p.m.
11	Mar.	Fri.	Spring recess begins, 6 p.m.

21	Mar.	Mon.	Spring recess ends; classes resume, 8:10 a.m.
25	Apr.	Mon.	Classes end, 9 p.m.
26	Apr.	Tues.	Reading Period begins.
1	May	Sun.	Reading Period ends.
2	May.	Mon.	Examination period begins, 8:30 a.m.
16	May	Mon.	Examination period ends; <u>spring term ends</u> , 5 p.m.
23	May	Mon.	UNIVERSITY COMMENCEMENT
1	June	Wed.	Degrees voted by faculty.
1	Aug.	Mon.	Fifth-term certification deadline for rising third-year students.

**Labor Day is on September 7, 2015. Rosh Hashanah is on September 14-15, 2015 (beginning at sundown on September 13). Yom Kippur is on September 23, 2015 (beginning at sundown on September 22).

** YLS Alumni Weekend is October 22-25.